





SOFTWARE: Patentin version 3.2  
 SEQ ID NO 306  
 LENGTH: 15234  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 US-11-136-527-306

Query Match 10.2%; Score 50.2; DB 14; Length 15234;  
 Best Local Similarity 49.8%; Pred. No. 0.0026;  
 Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

20 CGAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAATCCCGAGAGAGGCCCA 79  
 7947 CGAGCAGAGAGAGAGCCCAAGCTTGAGCAGCTTTTCCAGAGCAGAGTGGCAAAACCA 8006  
 80 GCTGTCCCTGCTGGAAGCCGAGGCTGGCTGTATGAGAGATCTTCTCCGAGGGTGC 139  
 8007 GCTGCAG 8066  
 140 TGCCTGTGTGTCAGAGAGAGCTCAGAGAGCTGAGAGAGTCTGGCGGCTTGAAGCTGCT 199  
 8067 GGCAG 8126  
 200 GGAAGAAAGTCTGTGAGCTTCATCAGAACTGGCATCTGCAGAGAGATGGAAGTTC 259  
 8127 GCAAG 8186  
 260 GGGGAGAGAAATGCT 274  
 8187 GGAGAACCAAGAGCT 8201

RESULT 5  
 US-11-121-086-5  
 Sequence 5, Application US/1121086  
 Publication No. US20050266459A1  
 GENERAL INFORMATION:  
 APPLICANT: POULSEN, TIM S.  
 APPLICANT: NIELSEN, KIRSTEN V.  
 TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
 FILE REFERENCE: 09138, 6000-00000  
 CURRENT APPLICATION NUMBER: US/11/121,086  
 PRIOR FILING DATE: 2005-05-04  
 PRIOR APPLICATION NUMBER: 60/567,570  
 NUMBER OF SEQ ID NOS: 107  
 SOFTWARE: Patentin version 3.3  
 SEQ ID NO 5  
 LENGTH: 153376  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-11-121-086-5

Query Match 9.8%; Score 48.2; DB 14; Length 153376;  
 Best Local Similarity 53.4%; Pred. No. 0.014;  
 Matches 101; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

2 GGGCGGGCGGGGAGATGCCAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAATTT 61  
 16201 GGAGAGAACAG 16260  
 62 CCCGAG 121  
 16261 ACAG 16320  
 122 GTCTTCTCCGAGAGGCTGTGCTGTGTGTCAGAGAGAGCTCAGAGAGTGTCTG 181  
 16321 GAG 16380  
 182 GCGGAGCTT 190  
 16381 GTGGGTCTT 16389

RESULT 6  
 US-10-330-773-631/c  
 Sequence 631, Application US/10330773  
 Publication No. US20060040262A1  
 GENERAL INFORMATION:  
 APPLICANT: David W. Morris  
 APPLICANT: Marc Malandro  
 TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
 FILE REFERENCE: 529452001300  
 CURRENT APPLICATION NUMBER: US/10/330,773  
 CURRENT FILING DATE: 2002-12-27  
 NUMBER OF SEQ ID NOS: 981  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 631  
 LENGTH: 30191  
 TYPE: DNA  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)...(30191)  
 OTHER INFORMATION: n = A,T,C or G  
 US-10-330-773-631

Query Match 9.5%; Score 46.6; DB 9; Length 30191;  
 Best Local Similarity 52.3%; Pred. No. 0.027;  
 Matches 103; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

9 CCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAATTCGCGAG 68  
 13754 CAGAGCAG 13695  
 69 AAGAGGCCCAAGCTGTCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128  
 13694 CAG 13635  
 129 CCGAGAGGTCCTCCCTGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188  
 13634 CAGAGCAG 13575  
 189 TTGAGGCTGTGAGAG 205  
 13574 CAGAGCTGAGAGAGAG 13558

RESULT 7  
 US-10-330-773-905  
 Sequence 905, Application US/10330773  
 Publication No. US20060040262A1  
 GENERAL INFORMATION:  
 APPLICANT: David W. Morris  
 APPLICANT: Marc Malandro  
 TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
 FILE REFERENCE: 529452001300  
 CURRENT APPLICATION NUMBER: US/10/330,773  
 CURRENT FILING DATE: 2002-12-27  
 NUMBER OF SEQ ID NOS: 981  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 905  
 LENGTH: 138808  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)...(138808)  
 OTHER INFORMATION: n = A,T,C or G  
 US-10-330-773-905

Query Match 9.2%; Score 45.2; DB 9; Length 138808;  
 Best Local Similarity 52.7%; Pred. No. 0.086;  
 Matches 96; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

30 GAGGCCAGTTTGAAGAGCTGTGTGCAAGATTTCCCGAGAGAGAGAGAGAGAGAGAG 89





```
; CURRENT APPLICATION NUMBER: US/11/036,196
; CURRENT FILING DATE: 2005-01-18
; PRIOR APPLICATION NUMBER: US/10/152,319
; PRIOR FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2106
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. U02506
US-11-036-196-2106
```

```
Query Match      8.5%; Score 42; DB 11; Length 648;
Best Local Similarity 47.7%; Pred. No. 0.22;
Matches 123; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
```

```
QY 47 GCTGTGGCGAGAAATTCGCCGAGAGAGGCCAGCTGTCCTCGTGGAGACCGAGGCGTG 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 GTTGTGTTCAAGACAGAGAGAGAGAGAGAGACAGAACAGAACAGAGAGAGCGAGGAGAA 154

QY 107 GCTGTGATGAGAAATCTTCTCCGAGAGGTGCTGCCGTGTCAGAGAGAGCTCAGGGA 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 GGAAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 214

QY 167 GCTGCGAGAGTGTGCGGCGCTTGAAGCTGTGAAAGAAAGTCTGAGCCTCATCAG 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 274

QY 227 AAACCTGCATCTGCAGAGATGGAATGATTGGGGAGAAAGTTTTCACCAACAA 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 GAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 334

QY 287 CATCCCAAGTCAGATT 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 GAAAGAAAGAGAAATT 352
```

```
RESULT 15
US-11-096-568A-21841/C
; Sequence 21841, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21841
; LENGTH: 1811
; TYPE: DNA
```

```
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1811)
; OTHER INFORMATION: Ceres Seq. ID no. 12406472
US-11-096-568A-21841
```

```
Query Match      8.5%; Score 41.8; DB 11; Length 1811;
Best Local Similarity 50.8%; Pred. No. 0.3;
Matches 100; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
```

```
QY 2 GCGCGCGCGGAGATGCCAGTCCCAAGAGCGCAGTTTGAGAGCTGTGGCAGAAAT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 GGCCTCTCTGACGGGTGCCGAGATCCATGGGCGCTTGATAGCGGCATGTGTGTGA 514

QY 62 CCGGAGAAAGAGGCCCAAGCTGTCCCTGTGTGAAGCCAGGCTGCTGTGTATGAGAA 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 513 GCGCGAGGCGCTCCACTGACGGGCGGCTTGAAACACACGCTGCTGTGTCTTGGCGCA 454

QY 122 GTCCTTCGCGAGGGGTGCTGCCGTGTGTCAGAGAGAGCTCAGGAGCTGGCAGAGTCGTG 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 GCTTCTGAGGATCTGGCCGACCGCTTCCGATGCCCGCGAGCACCGCGCGCGCG 394

QY 182 GCGGCTTGAAGCTGC 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 GCGGCGCTGCTGCTGC 377
```

```
Search completed: April 27, 2006, 05:41:19
Job time : 3289 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 03:03:54 ; Search time 3013 Seconds  
(without alignments)  
9300.970 Million cell updates/sec

Title: US-10-071-510A-16

Sequence: 1 cggccgcggccggggatgcc.....cctctctcatcttgagcctgc 493

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapexc 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: gb\_Da:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_com:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_to:\*  
10: gb\_strs:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vi:\*  
14: gb\_htg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	490.4	99.5	5502	8	HSM808450
2	347.6	70.5	1938	8	AK098471
3	347.6	70.5	3978	6	CQ735548
4	345.4	70.1	3003	6	CQ845922
5	345.4	70.1	3003	8	AK131436
6	283.4	57.5	150350	8	CNS01DMV
7	283.4	57.5	166007	14	CNS01DRC
8	212.2	43.0	3370	9	BC099694
9	140.6	28.5	165217	14	AC152545
10	121.4	24.6	176510	9	AC147375
11	57	11.6	110000	15	AP008207_101
12	57	11.6	142268	15	AP003018
13	57	11.6	154248	15	AP003631
14	57	11.6	159749	15	AP003020
15	56.8	11.5	2279	2	AY571308
16	56	11.4	110000	15	AP008207_037
17	56	11.4	154137	15	AP003215
18	55.4	11.2	110000	15	AP008214_114

C 19	55.4	11.2	135511	15	AP005097	AP005097 Oryza sat
C 20	55.4	11.2	154441	15	AP005383	AP005383 Oryza sat
C 21	55.2	11.2	194736	9	AC126271	AC126271 Mus muscu
C 22	54	11.0	109894	14	AP005975	AP005975 Oryza sat
C 23	54	11.0	110000	15	AP008213_014	Continuation (15 o
C 24	54	11.0	132526	15	AP006451	AP006451 Oryza sat
C 25	54	11.0	147782	15	AP005604	AP005604 Oryza sat
C 26	53.8	10.9	55528	2	AC105764	AC105764 Caenorhab
C 27	53.8	10.9	213476	14	AC114589	AC114589 Mus muscu
C 28	53.6	10.9	78835	15	AC120528	AC120528 Oryza sat
C 29	53.6	10.9	110000	15	AP008217_098	Continuation (99 o
C 30	53.4	10.8	110000	15	AP008213_013	Continuation (14 o
C 31	53.4	10.8	169511	15	AP005309	AP005309 Oryza sat
C 32	53.4	10.8	211542	15	AC126275	AC126275 Mus muscu
C 33	53	10.8	541	15	CR354530	CR354530 Plinu pin
C 34	53	10.8	110000	15	AP008207_416	Continuation (417
C 35	53	10.8	110000	15	AP008208_227	Continuation (228
C 36	53	10.8	128223	15	AP005643	AP005643 Oryza sat
C 37	53	10.8	176530	15	AP003259	AP003259 Oryza sat
C 38	52.8	10.7	125020	8	AF429315	AF429315 Homo sapi
C 39	52.8	10.7	217522	9	AC117639	AC117639 Mus muscu
C 40	52.8	10.7	227897	14	AC111962	AC111962 Rattus no
C 41	52.8	10.7	237997	14	AC093989	AC093989 Rattus no
C 42	52.6	10.7	253504	9	AC102602	AC102602 Mus muscu
C 43	52.2	10.6	97839	14	AC141520	AC141520 Rattus no
C 44	52.2	10.6	110000	15	AP008217_250	Continuation (251
C 45	52.2	10.6	110000	15	AP008208_094	Continuation (95 o

#### ALIGNMENTS

RESULT 1  
LOCUS HSM808450  
DEFINITION Homo sapiens mRNA; cDNA DKFZp686M07125 (from clone DKFZp686M07125).  
ACCESSION BX648302  
VERSION BX648302.1 GI:34367461  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
1 (bases 1 to 5502)  
Fboerker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,  
Rohde,G., Han,M. and Wiemann,S.  
The German Human cDNA Consortium  
Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuherberg, GERMANY  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
This clone (DKFZp686M07125) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
information about the clone and the sequencing project is available  
at http://mips.gsf.de/proj/cDNA/.

#### COMMENT

Location/Qualifiers  
1..5502  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp686M07125"  
/tissue\_type="human uterus"  
/clone\_id="686 (synonym: nlcc3). Vector pSPORT1\_Sfi; host  
DH10B; sites SfiI + SfiIB  
/dev\_stage="adult"  
5474

#### FEATURES

source  
polysite  
ORIGIN

Query Match	99.5%;	Score 490.4;	DB 8;	Length 5502;
Best Local Similarity	99.8%;	Pred. No. 1,4e-96;		
Matches 491;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	2	GGCGGCGCGGGGAGATGCGCAGTCCCAAGAGCGAGTTTGAGAGGCTGTGGCAGAAATT	61	
Db	3918	GGGGGGGCGGGGGGATGCCAGTCCCAAGAGCGCAGTTTGAGAGGCTGTGGCAGAAATT	3977	
QY	62	CCCGGAGAGAGAGAGCCCAAGCTGTCTCTGTGTGAAGCGCAGGCTGGCTGTGATGAGAA	121	
Db	3978	CCCGGAGAGAGAGAGCCCAAGCTGTCTCTGTGTGAAGCGCAGGCTGGCTGTGATGAGAA	4037	
QY	122	GTCTTTCCGCGAGGAGTGTGCGCGTGTGTCAGAGAGAGCTGAGGAGCTGGCAGAGTCGTG	181	
Db	4038	GTCTTTCCGCGAGGAGTGTGCGCGTGTGTCAGAGAGAGCTGAGGAGCTGGCAGAGTCGTG	4097	
QY	182	GGCGGCGCTTGAGAGCTGCTGGAAGAAAGTCTGTGAGCCTCATGAAACTGGCATCTGCA	241	
Db	4098	GGCGGCGCTTGAGAGCTGCTGGAAGAAAGTCTGTGAGCCTCATGAAACTGGCATCTGCA	4157	
QY	242	GAGGATGGAAGTGGATTTGGGGAGAAATGTTTTCACCAACAACATCCCAAAATCAGG	301	
Db	4158	GAGGATGGAAGTGGATTTGGGGAGAAATGTTTTCACCAACAACATCCCAAAATCAGG	4217	
QY	302	ATTCTTCATCAATCCCATGATGATCTCTATTCACAGCATCTGTGAGCGGTGATGCTGTAG	361	
Db	4218	ATTCTTCATCAATCCCATGATGATCTCTATTCACAGCATCTGTGAGCGGTGATGCTGTAG	4277	
QY	362	CAGGCGCTGTGGAGAGAGGCGCCAGGCCCAAGTCAAGAGGTGGGTAGGGGCTTCCAGCAC	421	
Db	4278	CAGGCGCTGTGGAGAGAGGCGCCAGGCCCAAGTCAAGAGGTGGGTAGGGGCTTCCAGCAC	4337	
QY	422	AGGCGCTCTCTCTGTCTGTGGGCAACATGCTCTGCTCTGAGAGACTTTGGCACGTCCTGTCTC	481	
Db	4338	AGGCGCTCTCTCTGTCTGTGGGCAACATGCTCTGCTCTGAGAGACTTTGGCACGTCCTGTCTC	4397	
QY	482	ATTGAGCGCTGC 493		
Db	4398	ATTGAGCGCTGC 4409		
RESULT 2				
AK098471		1938 bp	mRNA	linear
LOCUS	AK098471			
DEFINITION	Homo sapiens cDNA FLJ25605 f18, clone JTH14272.			
ACCESSION	AK098471			
VERSION	AK098471.1 GI:21758489			
KEYWORDS	oligo capping; fig (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
	Homidae; Homo.			
REFERENCE	1			
AUTHORS	Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,			
	Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,			
	Kamihara, K., Katubata, N., Sato, K., Tanikawa, M., Yamazaki, M.,			
	Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S., Morinaga, M.,			
	Kawanura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,			
	Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isegai, T. and			
	Sugano, S.			
	NEDO human cDNA sequencing project			
TITLE	Unpublished			
JOURNAL	2 (bases 1 to 1938)			
REFERENCE	Sugano, S. and Suzuki, Y.			
AUTHORS	Direct Submission			
TITLE	Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,			
JOURNAL	University of Tokyo, Laboratory of Genome Structure, Human Genome			
	Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan			
	(E-mail: flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,			
	Fax:81-3-5449-5416)			
COMMENT	NEDO human cDNA sequencing project supported by Ministry of			

Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction and 5'-end one pass sequencing; Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing; RAB; clone selection for full insert sequencing; RAB and Helix Research Institute.

FEATURES	
source	1. 1938
location/Qualifiers	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="JTH14272"
	/cell_line="JCR"
	/tissue_type="thyroid"
	/clone_id="JTH"
	/note="cloning vector: pME18FLJ3"
CDS	237..1706
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="BAC05312.1"
	/db_xref="GI:21758490"
	/translation="MESSRLKELLTMDLQKDLIGTQGERATLLEQVAGSMRDRLHNSLQKSKLQSLAQHDFGAFFPLQKLDLQVVAQEKGLQDLDPKQAOQLSRLOGLQEBGLDIGHQMEAPRLVQENPHQMDLSDFOALQSLDLDVRCQSSVDEHTFSPHQLLELROMIVVTTQKLEHAGEAPGDAESQEAFFERLVAEFPKXAOVSLVEAQGLVMEKSPREGAAVVOEELRELAEKSWRLRLLESLSLINMHQRMEDVSGKRVFPTNNIPKSGFLINPMDIPRHRBRANTLOEBSHSDPSQLRNFQOMLOEYNSKLVRIITAMRTSTADLRTKRSKLDLELRVPEQGHLENILRIGPARGTDEDELRYOMLVYSKLDGSHLLTQSSPGEPTGQKTRNRGUGSLFRVCCVALPLQLLLFLFLFLPIREDRSLTANNFARSLTLMRYNPPPT"
ORIGIN	
Query Match	70.5%; Score 347.6; DB 8; Length 1938;
Best Local Similarity	98.9%; Pred. No. 2e-65;
Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	2 GGCGGCGCGGGGAGATGCGCAGTCCCAAGAGCGCAGTTTGAGAGGCTGTGGCAGAAATT 61
Db	809 GGCGGCGCGGGGAGATGCGCAGTCCCAAGAGCGCAGTTTGAGAGGCTGTGGCAGAAATT 868
QY	62 CCCGAGAAAGAGGCCAGCTGTCTCTGTGTGAAGCGCAGGCTGGCTGTGATGAGAA 121
Db	869 CCCGAGAAAGAGGCCAGCTGTCTCTGTGTGAAGCGCAGGCTGGCTGTGATGAGAA 928
QY	122 GTCTTCTCCGAGAGGCTGCTCCGTGTGTGACAGAGAGCTCAGAGAGCTGGCAGAGTCGTG 181
Db	929 GTCTTCTCCGAGAGGCTGCTCCGTGTGTGACAGAGAGCTCAGAGAGCTGGCAGAGTCGTG 988
QY	182 GGCGGCGCTTGAGGCTGTCTGGAAGAAAGTCTGTGAGCCTCATCAGAAACTGGCATCTGCA 241
Db	989 GGCGGCGCTTGAGGCTGTCTGGAAGAAAGTCTGTGAGCCTCATCAGAAACTGGCATCTGCA 1048
QY	242 GAGGATGGAAGTGGATTTGGGGAGAAATGTTTTCACCAACAACATCCCAAAGTCAGG 301
Db	1049 GAGGATGGAAGTGGATTTGGGGAGAAATGTTTTCACCAACAACATCCCAAAGTCAGG 1108
QY	302 ATTTCATCAATCCCATGATGATCTCTATTCACAGCATCTGTGAGCGGTGATGCTGTAG 355
Db	1109 ATTTCATCAATCCCATGATGATCTCTATTCACAGCATCTGTGAGCGGTGATGCTGTAG 1162
RESULT 3	
CQ735548	3978 bp DNA linear PAT 03-FEB-2004
LOCUS	CQ735548
DEFINITION	Sequence 21482 from Patent WO02068579.
ACCESSION	CQ735548
VERSION	CQ735548.1 GI:42327813
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.



REFERENCE 1  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE K15, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 21482 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES location/Qualifiers  
source 1..3978  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 70.5%; Score 347.6; DB 6; Length 3978;  
Best Local Similarity 98.9%; Pred. No. 1.9e-65;  
Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGGCGGCGCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGCAGAAATT 61  
DB 2847 GGGCGGCGCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGCAGAAATT 2906  
QY 62 CCGGAGAGAGAGGCGCCAGCTGTCTCTGTGTGAAGCCGAGGCTGTGTGTGATGAGAA 121  
DB 2907 CCGGAGAGAGAGGCGCCAGCTGTCTCTGTGTGAAGCCGAGGCTGTGTGTGATGAGAA 2966  
QY 122 GTCTTCTCCGAGAGGCTGTCTCTGTGTGAAGAGAGAGCTCAGGAGCTGTGAGAGTGTG 181  
DB 2967 GTCTTCTCCGAGAGGCTGTCTCTGTGTGAAGAGAGAGCTCAGGAGCTGTGAGAGTGTG 3026  
QY 182 GCGGCGCTTGAGAGCTGTGTGAAGAAAGTCTGTGAGCTCATCGAAATCTGCA 241  
DB 3027 GCGGCGCTTGAGAGCTGTGTGAAGAAAGTCTGTGAGCTCATCGAAATCTGCA 3086  
QY 242 GAGGATGGAAGTGAATTCGGGGAAGAAATGTTTTCACCAACAATCCCAAGTCAAG 301  
DB 3087 GAGGATGGAAGTGAATTCGGGGAAGAAATGTTTTCACCAACAATCCCAAGTCAAG 3146  
QY 302 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGCTGAGTCT 355  
DB 3147 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGCGCAATCT 3200

RESULT 4  
LOCUS CQ845922 3003 bp DNA linear PAT 02-AUG-2004  
DEFINITION Sequence 4569 from Patent EP1440981.  
ACCESSION CQ845922  
VERSION CQ845922.1 GI:50895521  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Iehii, S., Yamamoto, J., Isono, Y., Nagai, K. and Irie, R.  
TITLE Full-length human cdna  
JOURNAL Patent: EP 1440981-A 4569 28-JUL-2004;  
Research Association for Biotechnology (JP)  
FEATURES location/Qualifiers  
source 1..3003  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 70.1%; Score 345.4; DB 6; Length 3003;  
Best Local Similarity 99.7%; Pred. No. 5.9e-65;  
Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGCGGCGCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGCAGAAATT 61

DB 1911 GGGCGGCGCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGCAGAAATT 1970  
QY 62 CCGGAGAGAGAGGCGCCAGCTGTCTCTGTGTGAAGCGCAGGCTGTGTGTGATGAGAA 121  
DB 1971 CCGGAGAGAGAGGCGCCAGCTGTCTCTGTGTGAAGCGCAGGCTGTGTGTGATGAGAA 2030  
QY 122 GTCTTCTCCGAGAGGCTGTCTCTGTGTGAAGAGAGCTCAGGAGCTGTGAGAGTGTG 181  
DB 2031 GTCTTCTCCGAGAGGCTGTCTCTGTGTGAAGAGAGCTCAGGAGCTGTGAGAGTGTG 2090  
QY 182 GCGGCGCTTGAGAGCTGTGTGAAGAAAGTCTGTGAGAGCTCATCGAAATCTGCA 241  
DB 2091 GCGGCGCTTGAGAGCTGTGTGAAGAAAGTCTGTGAGAGCTCATCGAAATCTGCA 2150  
QY 242 GAGGATGGAAGTGAATTCGGGGAAGAAATGTTTTCACCAACAATCCCAAGTCAAG 301  
DB 2151 GAGGATGGAAGTGAATTCGGGGAAGAAATGTTTTCACCAACAATCCCAAGTCAAG 2210  
QY 302 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGG 348  
DB 2211 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGG 2257

RESULT 5  
LOCUS AK131436 3003 bp mRNA linear PRI 07-MAY-2004  
DEFINITION Homo sapiens cDNA FLJ16564 f1s, clone SYN04005889.  
ACCESSION AK131436  
VERSION AK131436.1 GI:4707387  
KEYWORDS oligo capping, f1s (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Oca, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Iehii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsu, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isegai, T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3003)  
AUTHORS Isegai, T. and Yamamoto, J.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2004) Takao Isegai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES location/Qualifiers  
source 1..3003  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="SYNOV4005889"  
/tissue type="synovial membrane tissue from rheumatoid arthritis"  
/clone\_id="SYNOV4"  
/note="cloning vector: PME18SFL3"  
79..2793  
/note="unamed protein product"  
/codon\_start=1



```

QY      373 GAGAAAGGGGCGAGCCCGCCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 432
Db      90770 GAGAAAGGGGCGAGCCCGCCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 90711
QY      433 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTGGCCACGCTCTGTCTCACTTTGAGCCTG 492
Db      90710 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTGGCCACGCTCTGTCTCACTTTGAGCCTG 90651
QY      493 C 493
Db      90650 C 90650

RESULT 7
LOCUS   CNS01DRC 166007 bp DNA linear HTG 10-MAR-2005
DEFINITION Homo sapiens chromosome 14 clone R-298123, WORKING DRAFT SEQUENCE.
ACCESSION AL117186
VERSION   AL117186.3 GI:7710964
KEYWORDS  HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homiidae; Homo.
          1 (bases 1 to 166007)
REFERENCE Genoscope.
          Direct Submission
          Submitted (23-MAY-2000)
          On May 5, 2000 this sequence version replaced gi:6981774.
          IMPORTANT: This sequence is unfinished and does not necessarily
          represent the correct sequence. Work on the sequence is in progress
          and the release of this data is based on the understanding that the
          sequence may change as work continue. The sequence may be
          contaminated with foreign sequence from E. coli, yeast, vector,
          plasmid, etc. . . even if efforts are made to eliminate these
          contaminating sequences. The following BAC sequence is oriented
          from the T7 to the SP6 end.
          Upstream BAC (overlapping the T7 end) : R-1070N10
          Downstream BAC (overlapping the SP6 end) : C-2588C21
          -----
          Overall quality chart :
          Range : bases
          0 : 4
          1 - 9 : 676
          10 - 19 : 1988
          20 - 29 : 5400
          30 - 39 : 19159
          40 - 49 : 16768
          50 - 59 : 13248
          60 - 69 : 21617
          70 - 79 : 42283
          80 - 89 : 30090
          90 - 99 : 14774
          -----
          Percentage of bases with a quality value >= 40 : 83 %.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submitter.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          * 1 166007: contig of 166007 bp in length.
          Location/Qualifiers
            1..166007
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="14"
              /clone="R-298123"
FEATURES
SOURCE

```

```

ORIGIN
Query Match 57.5%; Score 283.4; DB 14; Length 166007;
Best Local Similarity 96.3%; Pred. No. 1,4e-51;
Matches 290; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

/cclone_lib="RPCL-11"

QY      193 GGGCTCTGGAAGAAATGCTGTGAGCTTCATCAGAAATCGGATTCGACAGAGATGAAG 252
Db      92267 GACTGGCTGAGATTGTCTCTTCACGCTCATAGAAATCGGATTCGACAGAGATGAAG 92326
QY      253 TGGATTGGGGGAGAAATGTTTTTTCACCAACAATCCCAAGATTCAGATTTCATCA 312
Db      92327 TGGATTGGGGGAGAAATGTTTTTTCACCAACAATCCCAAGATTCAGATTTCATCA 92386
QY      313 ATCCCATGATCTTATTCATCCAGCATGTCGACGCGTGAAGTCTGTACAGAGGCTGTG 372
Db      92387 ATCCCATGATCTTATTCATCCAGCATGTCGACGCGTGAAGTCTGTACAGAGGCTGTG 92446
QY      373 GAGAAAGGGGCGAGCCCGCCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 432
Db      92447 GAGAAAGGGGCGAGCCCGCCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 92506
QY      433 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTGGCCACGCTCTGTCTCACTTTGAGCCTG 492
Db      92507 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTGGCCACGCTCTGTCTCACTTTGAGCCTG 92566
QY      493 C 493
Db      92567 C 92567

RESULT 8
LOCUS   BC099694 3370 bp mRNA linear ROD 25-JUN-2005
DEFINITION Mus musculus cDNA clone MGC:106312 IMAGE:4507746, complete cds.
ACCESSION BC099694
VERSION   BC099694.1 GI:71043409
KEYWORDS  MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Muridae; Murinae; Mus.
          1 (bases 1 to 3370)
REFERENCE Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
          Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
          Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
          Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
          Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
          Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
          Scheetz, T.E., Brownstein, M.J., Uedl, T.B., Toshiyuki, S.,
          Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
          Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
          McKernan, K.O., Malek, J.A., Gunaratne, P.H., Richards, S.,
          Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
          Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
          Fahey, J.J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S.,
          Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
          Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
          Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
          Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,
          Schnerbach, A., Schein, J.E., Jones, S.J. and Marra, M.A.
          Mammalian Gene Collection Program Team
          Generation and initial analysis of more than 15,000 full-length
          human and mouse cDNA sequences
          Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)
          12477932
          2 (bases 1 to 3370)
          NIH MGC Project
          Direct Submission
          Submitted (15-JUN-2005) National Institutes of Health, Mammalian

```

REMARK  
COMMENT

Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaitheerburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@hgrl.nih.gov](mailto:nisc.mgc@hgrl.nih.gov)  
Akhter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legault, R.,  
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W.,  
Teague, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAC Plate: 208 Row: d Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 27359681.

## FEATURES

## source

## CDS

1..3370  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="MGC:106312 IMAGE:4507746"  
/tissue\_type="eye, retina, mouse strain C57BL/6"  
/clone\_lib="NIH\_MGC\_94"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
293..3220  
/codon\_start=1  
/product=Unknown (protein for MGC:106312)"  
/protein\_id="AAH9694.1"  
/db\_xref="GI:71043410"  
/translation="MTQGPDEPERSVEDAQAAMKVIYIOLQVNDNTKPRALAEAL  
RETERKICOLEBGMVLELIRAAALATCOGKPEILATGKQKMESEVYVLT  
HCHSRTEWVHNSVLELADDERPFRQKVVAVLPPVLOGLDKKQKSHAVLT  
HNDVNAVLDRLLEBAGSLFSGIPSDVEDQKMKABYDVAKRARRDVLQV  
ADHEQVREDVNEFQMLKAVKHSCLGRNCKLATELRLSTLDIADPFGESL  
KRLEQAVGVONTSPLEAKESLSELEMGVLEKLVLMKEBGRGLQSGDCE  
OOIQOLEAEIGDFKSLQRLAORGLPEPTVATEDELVAQMLFSGTBAALASEPRV  
DRIQOLKLVLPDLOSLSDSVATIOEQSMKGNTRLHNTAELMORFORPLND  
LQMKRLAQLDLDTASLPLASITHTFLPQIAALRESSRLKQMLADLKTDLGSI  
FQGERATLLEQVTSVRDLDLHNSLQKSLQSLVLOHNDFGAFLPKRLDL  
QARIQAEKGLPDLPGKQVQLRLQGLQEGDLQGIQIAVPRPLAAGSKHQKQVQI  
SCQOALQSLLELDVRCQNVNEHCTFSHRLSELQMTMTAQITLESQGVKMDA  
ESQEALETLLSEIPEKEVQSLQALQGLVMKSSPEGAATVQELERKLSMWQRL  
LLEENMLSLMRNQLORTQVDTGKQVFNINIPKAGFLINPDPIPRQHGANPLEGH  
DLPEHDLPQLRDEPQMLQANSLKRLIITVRVATADLPTREYKQLEEARLPEQGH  
FEKLRLRPADPSNELEDRTIKMPLFKSLQSGHLLTBSSGELTARQKSRQKRW  
SPSLIQKCRVALPQLLLFLFLFLPLPAGEEBSICALANNFARSPALMLRYNGP  
PPT"

## ORIGIN

Query Match 43.0%; Score 212.2; DB 9; Length 3370;  
Best Local Similarity 77.8%; Pred. No. 6.1e-36;  
Matches 256; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 7 GCCCGGGGATGCGAGTCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGATTCCTCCG 66  
DB 2325 GTCCTGGAGATGTGAGTCCCAAGAGCTGAGTCCGAGAGCTGCTCTCTGGAATCCAG 2384  
QY 67 AGAAGAGGCCAGCTGTCCCTGTGAGAGCGAGGCTGGCTGTGTGATGAGAGACTCTT 126  
DB 2385 AGAAAGAGGTCAGAGTGTCTCTCAAGACATGGGCGAGCTTGTGATGAGAGACTT 2444

QY 127 CTCGGAGAGGTCTGCGCGTGTGACAGAGAGAGCTCAGGAGCTGACAGAGTCTGCGCG 186  
DB 2445 CCCCAAGAGGGCAACCATGTGTCAGAGAGAGCTAGAGAGTGTGAGTCTTGGCAGG 2504  
QY 187 CCTTGAAGGTCTGTGAGAGAAAGTCTGTGAGCTTGTGAGAGAACTGGATCTGACAGGA 246  
DB 2505 CCTCGGCGGTGTGAGAGAGAAACATGTGAGTGTGATGAGAGAAACAGAGCTGACAGGA 2564  
QY 247 TGGAGTGTGATTTGGGGGAAAGAAATGCTTTTACCAACAACATCCCAAGTACAGATTTC 306  
DB 2565 CAGAGGTGACACCGGGGAAAGAGAGGTTCACCAACAACATCCCAAGCGGCTTTTC 2624  
QY 307 TCATCAATCCATGGATCTATTCACAG 335  
DB 2625 TCATCAATCCATGGATCTATTCACAG 2653

REFERENCE  
AUTHORS

RESULT 9  
AC152545/C  
LOCUS  
DEFINITION  
AC152545 165217 bp DNA linear HTG 01-JUL-2005  
Bos taurus clone CH240-3M23, WORKING DRAFT SEQUENCE, 16 unordered  
pieces.  
AC152545  
AC152545.4 GI:68227057  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
Bos taurus (cow)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 165217)  
Muzny, D., Marie, M., Metker, M., Lee, J., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D.,  
Anyalele, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandarama, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenship, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Devila, M., Davis, C., Davy-Carroll, J., De And, C., Dederich, D.,  
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guayara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lounsbury, L., Louised, H., Lozdo, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, B.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munda, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwackeme, O., Okunnu, G., Olarunpata, A., Pal, S., Parke, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L., L.,  
Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smaj, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K.,

TITLE	Journal Unpublished
JOURNAL TITLE	(bases 1 to 165217)
AUTHORS	Morley, K.C.
JOURNAL	Submitted (12-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 165217) Cow Genome Sequencing Consortium. Direct Submission Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
AUTHORS	Valas, R., Veara, V., Villaseña, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Morley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O., Meisner-Gugan, G. and Gibbs, R.A.
JOURNAL	Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	On Jun 26, 2005 this sequence version replaced gi:58037954. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
JOURNAL	Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu Project information Center project name: FAMJW Center clone name: CH240-3M23 Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 155449 bases at least Q40 Consensus quality: 156707 bases at least Q30 Consensus quality: 157568 bases at least Q20 Estimated insert size: 157550; sum-of-contigs estimation Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
REFERENCE	* NOTE: Estimated insert size may differ from sequence length (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html</a> ) * NOT: This is a "working draft" sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
JOURNAL	1 33332: contig of 33332 bp in length 33382: gap of 50 bp 33383: 46869: contig of 13487 bp in length 46870: 46919: gap of 50 bp 46920: 54137: contig of 7228 bp in length 54138: 54695: gap of 356 bp 54696: 72807: contig of 18112 bp in length 72808: 74316: gap of 1509 bp 74317: 76686: contig of 2370 bp in length 76687: 76765: gap of 79 bp 76766: 89950: contig of 13385 bp in length 89951: 91560: gap of 1610 bp 91561: 110253: contig of 18693 bp in length

FEATURES	SOURCE
110254	110303: gap of 50 bp
110304	120118: contig of 9815 bp in length
120119	120554: gap of 436 bp
120555	121919: contig of 11365 bp in length
131920	131969: gap of 50 bp
131970	136597: contig of 4628 bp in length
136598	136697: gap of unknown length
136698	138982: contig of 2285 bp in length
138983	139037: gap of 50 bp
139033	147525: contig of 8493 bp in length
147526	148107: gap of 582 bp
148108	149195: contig of 1088 bp in length
149196	149295: gap of unknown length
149296	161575: contig of 12280 bp in length
161576	163037: gap of 1462 bp
163038	164066: contig of 1029 bp in length
164067	164166: gap of unknown length
164167	165217: contig of 1051 bp in length.
Location/Qualifiers	
1.	165217
/organism="Bos taurus"	
/mol_type="genomic DNA"	
/db_xref="taxon:9913"	
/clone="CH240-3M23"	
33333.	33382
/estimated_length=50	
46870.	46919
/estimated_length=50	
54138.	54695
/estimated_length=558	
72808.	74316
/estimated_length=1509	
76687.	76765
/estimated_length=79	
89951.	91560
/estimated_length=1610	
110254.	110303
/estimated_length=50	
120119.	120554
/estimated_length=436	
131920.	131969
/estimated_length=50	
136598.	136697
/estimated_length=unknown	
138983.	139032
/estimated_length=50	
147526.	148107
/estimated_length=582	
149196.	149295
/estimated_length=unknown	
161576.	163037
/estimated_length=1462	
164067.	164166
/estimated_length=unknown	
ORIGIN	
Query Match	28.5%; Score 140.6; DB 14; Length 165217;
Best Local Similarity	72.7%; Pred. No. 1.8e-20;
Matches 210; Conservative	0; Mismatches 74; Indels 5; Gaps 2;
QY	208 GTCGCTAGCCTCATCAGAACTGCGCATCTGCGAGGATG---GAAGTGGATTCCGGGA 264
DB	138160 GACTTCTAGCCTCTCCAGAAACCGGACGCTCAGAGATGACAGAGTGGATTCCAGGGA 1381010
QY	265 AGAAATGTTTTCACCAACACATCCCAAGTCAGATTTCTCATCAATCCCATGGATC 324
DB	138100 AGAAATGATTTTTCACCAACACATCCCAAGATCGGGTTTCTCATCTCCCAAGGATC 1380411
QY	325 CTATTCCAGGCGATGTGACCGCGTGAAGTCTGTCTTAGCAGGGGCTGTGGAGAAAGGGCCA 384
DB	138040 TTACTTCCCGGCGATCGCGTCAAGTAAGTCATCTTCATGCGGCTTCTGAGAGGGCCAGA 1379811
QY	385 GGCCCCAGGTCAAGAGTGGGTGAGGGGTCTCCAGACAGAGCCCTCCCTGTGTGGGGCAA 444

Db 137980 TCCGAGCTCAGTTGATGCTGGGAAGGTGTACCCAGCAAAGCCCCCTCCCATCTCGGGGCCA 1379221

Qy 445 CATGCTCTGCTCTGAGGACTTGAGCCACGTCCTGTCTCATTTGAGCCTGC 493

Db 137920 CATGCTCAGCTC--AGGACTGGGGCCCTGTCTCTGTCTGGGTGAGTCTGC 137874

RESULT 10  
AC147375 176510 bp DNA linear ROD 29-MAY-2004  
LOCUS AC147375/c  
DEFINITION Mus musculus BAC clone RP24-159013 from chromosome 12, complete  
sequence.  
VERSION AC147375  
KEYWORDS AC147375.3 GI:47564188  
SOURCE HTG.  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 176510)  
Levy, A. and Kozlowski, A.  
The sequence of Mus musculus BAC clone RP24-159013  
Unpublished (2001)  
2 (bases 1 to 176510)  
Wilson, R.K.  
Direct Submission  
Submitted (11-NOV-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 176510)  
Wilson, R.K.  
Direct Submission  
Submitted (17-DEC-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 176510)  
Wilson, R.K.  
Direct Submission  
Submitted (22-MAY-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 176510)  
Wilson, R.K.  
Direct Submission  
Submitted (29-MAY-2004) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On May 22, 2004 this sequence version replaced gi:3979582.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: M\_BB0159013  
-----

NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., paired quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCT-22BAC library has been constructed by Pleier de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pleier de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone. This clone is overlapped by AC133077 and AC140477.

FEATURES	
source	location/Qualifiers
	1. .176510
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="12"
	/map="12"
	/clone="RP24-159013"
	/clone_lib="RPC1-24"
repeat_region	813. .964
	/rpt_family="MIR"
repeat_region	1144. .1387
	/rpt_family="B4"
repeat_region	1394. .1716
	/rpt_family="MaLR"
repeat_region	1722. .2860
	/rpt_family="MaLR"
repeat_region	2861. .3004
	/rpt_family="A1u"
repeat_region	3005. .3117
	/rpt_family="MaLR"
repeat_region	3123. .3536
	/rpt_family="MaLR"
repeat_region	4400. .4646
	/rpt_family="B4"
repeat_region	5441. .5554
	/rpt_family="MER1_type"
repeat_region	5674. .5732
	/rpt_family="RMR1B"
repeat_region	5733. .5918
	/rpt_family="B2"
repeat_region	5919. .5937
	/rpt_family="RMR1B"
repeat_region	6186. .6331
	/rpt_family="A1u"
repeat_region	6461. .6507
	/rpt_family="MIR"
repeat_region	7024. .7144
	/rpt_family="MIR"
repeat_region	7619. .7686
	/rpt_family="B4"
repeat_region	7870. .8087
	/rpt_family="B4"
repeat_region	9190. .9254
	/rpt_family="MIR"
repeat_region	10468. .10828
	/rpt_family="L1"
repeat_region	11586. .11664
	/rpt_family="ERVX"
repeat_region	12182. .12555
	/rpt_family="L1"
repeat_region	13446. .13458
	/rpt_family="MaLR"
repeat_region	13459. .13537
	/rpt_family="A1u"
repeat_region	13538. .13562
	/rpt_family="MaLR"
repeat_region	13678. .13852
	/rpt_family="B2"
repeat_region	13894. .13938
	/rpt_family="MER1_type"
repeat_region	14211. .14367
	/rpt_family="B4"
repeat_region	14431. .14650
	/rpt_family="B2"

```
repeat_region      14666. .14853  
/rpt_family="B2"  
repeat_region      15844. .16144  
/rpt_family="B4"  
tRNA  
repeat_region      15847. .15917  
/product="tRNA-Va1"  
repeat_region      16202. .16286  
/rpt_family="MIR"  
repeat_region      18142. .18324  
/rpt_family="B2"  
repeat_region      18744. .18997  
/rpt_family="B4"  
repeat_region      19955. .20107  
/rpt_family="B4"  
repeat_region      20367. .20381  
/rpt_family="Alu"  
repeat_region      20382. .20530  
/rpt_family="Alu"  
repeat_region      20531. .20641  
/rpt_family="Alu"  
repeat_region      21096. .21381  
/rpt_family="B4"  
repeat_region      22347. .22445  
/rpt_family="L1"  
repeat_region      22472. .22533  
/rpt_family="ERV1"  
repeat_region      24262. .24977  
/rpt_family="L1"  
repeat_region      25408. .25593  
/rpt_family="B2"  
repeat_region      25605. .25680  
/rpt_family="B4"  
repeat_region      25718. .25792  
/rpt_family="ID"  
repeat_region      26024. .26158  
/rpt_family="MER1_type"  
repeat_region      27153. .27297  
/rpt_family="MER1_type"  
repeat_region      28246. .28543  
/rpt_family="MER2_type"  
repeat_region      29941. .30039  
/rpt_family="Alu"  
repeat_region      30808. .31101  
/rpt_family="B4"  
repeat_region      31418. .31467  
/rpt_family="ERV1"  
repeat_region      31652. .31777  
/rpt_family="MALR"  
repeat_region      32253. .32353  
/rpt_family="ACHobo"  
repeat_region      33136. .33381  
/rpt_family="MIR"  
repeat_region      33617. .33662  
/rpt_family="Alu"  
repeat_region      33711. .33763  
/rpt_family="ID"  
repeat_region      34334. .34710  
/rpt_family="MALR"  
repeat_region      36098. .36315  
/rpt_family="B4"  
repeat_region      36650. .36735  
/rpt_family="MIR"  
repeat_region      39916. .40251  
/rpt_family="RMER1B"
```

Query Match 24.6%; Score 121.4; DB 9; Length 176510;  
Best Local Similarity 69.8%; Pred. No. 2.8e-16;  
Matches 178; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 208 GCTCTGAGCTCTCATCAAGAACTGCACTGACAGATGGAAGTTCGGGAGGA 267  
DB 160158 GCTCCCACTCATGAGAAACGACGCTGACAGAGACAGAGTGCACAGGGGAGGA 160099

```
QY 268 AATGGTTTCCACCAACATCCCAAGTCAGATTTCATCATCCATCCATGATCTTA 327  
DB 160098 AGCAGGTTCACCAACATCCCAAGCCGCTTTCATCAACCTCAGAGCCCA 160039  
QY 328 TTCCAGGATCGTCAGCCGTAGTGTCTGTAGCAGAGGCTGTGGAGAGGGCCAGGC 387  
DB 160038 TTCCAGGATCGTCAGCCGTAGTGTCTGTAGCAGAGGCTGTGGAGAGGGCCAGGC 159980  
QY 388 CCCAGGTCAAGAGTGGTGGGTTCCAGACAGAGCCCTCCCTGTCTGGGGCAACAT 447  
DB 159979 AATGTCAAGTGAAGAGTGGGGCTCCAGCACAGTTCCTTCGCTGATCGCAT 159920  
QY 448 GCTTGTCTGAGGA 462  
DB 159919 GCTCAGCTCATGGA 159905  
  
RESULT 11  
AP008207_101/c  
WPCOMMENT  
Sequence split into 433 fragments LOCUS AP008207 Accession AP008207  
Fragment Name Begin End  
AP008207_000 1 11000  
AP008207_001 100001 210000  
AP008207_002 200001 310000  
AP008207_003 300001 410000  
AP008207_004 400001 510000  
AP008207_005 500001 610000  
AP008207_006 600001 710000  
AP008207_007 700001 810000  
AP008207_008 800001 910000  
AP008207_009 900001 1010000  
AP008207_010 1000001 1110000  
AP008207_011 1100001 1210000  
AP008207_012 1200001 1310000  
AP008207_013 1300001 1410000  
AP008207_014 1400001 1510000  
AP008207_015 1500001 1610000  
AP008207_016 1600001 1710000  
AP008207_017 1700001 1810000  
AP008207_018 1800001 1910000  
AP008207_019 1900001 2010000  
AP008207_020 2000001 2110000  
AP008207_021 2100001 2210000  
AP008207_022 2200001 2310000  
AP008207_023 2300001 2410000  
AP008207_024 2400001 2510000  
AP008207_025 2500001 2610000  
AP008207_026 2600001 2710000  
AP008207_027 2700001 2810000  
AP008207_028 2800001 2910000  
AP008207_029 2900001 3010000  
AP008207_030 3000001 3110000  
AP008207_031 3100001 3210000  
AP008207_032 3200001 3310000  
AP008207_033 3300001 3410000  
AP008207_034 3400001 3510000  
AP008207_035 3500001 3610000  
AP008207_036 3600001 3710000  
AP008207_037 3700001 3810000  
AP008207_038 3800001 3910000  
AP008207_039 3900001 4010000  
AP008207_040 4000001 4110000  
AP008207_041 4100001 4210000  
AP008207_042 4200001 4310000  
AP008207_043 4300001 4410000  
AP008207_044 4400001 4510000  
AP008207_045 4500001 4610000  
AP008207_046 4600001 4710000  
AP008207_047 4700001 4810000  
AP008207_048 4800001 4910000  
AP008207_049 4900001 5010000  
AP008207_050 5000001 5110000
```

AP008207\_051 5100001 5210000  
AP008207\_052 5200001 5310000  
AP008207\_053 5300001 5410000  
AP008207\_054 5400001 5510000  
AP008207\_055 5500001 5610000  
AP008207\_056 5600001 5710000  
AP008207\_057 5700001 5810000  
AP008207\_058 5800001 5910000  
AP008207\_059 5900001 6010000  
AP008207\_060 6000001 6110000  
AP008207\_061 6100001 6210000  
AP008207\_062 6200001 6310000  
AP008207\_063 6300001 6410000  
AP008207\_064 6400001 6510000  
AP008207\_065 6500001 6610000  
AP008207\_066 6600001 6710000  
AP008207\_067 6700001 6810000  
AP008207\_068 6800001 6910000  
AP008207\_069 6900001 7010000  
AP008207\_070 7000001 7110000  
AP008207\_071 7100001 7210000  
AP008207\_072 7200001 7310000  
AP008207\_073 7300001 7410000  
AP008207\_074 7400001 7510000  
AP008207\_075 7500001 7610000  
AP008207\_076 7600001 7710000  
AP008207\_077 7700001 7810000  
AP008207\_078 7800001 7910000  
AP008207\_079 7900001 8010000  
AP008207\_080 8000001 8110000  
AP008207\_081 8100001 8210000  
AP008207\_082 8200001 8310000  
AP008207\_083 8300001 8410000  
AP008207\_084 8400001 8510000  
AP008207\_085 8500001 8610000  
AP008207\_086 8600001 8710000  
AP008207\_087 8700001 8810000  
AP008207\_088 8800001 8910000  
AP008207\_089 8900001 9010000  
AP008207\_090 9000001 9110000  
AP008207\_091 9100001 9210000  
AP008207\_092 9200001 9310000  
AP008207\_093 9300001 9410000  
AP008207\_094 9400001 9510000  
AP008207\_095 9500001 9610000  
AP008207\_096 9600001 9710000  
AP008207\_097 9700001 9810000  
AP008207\_098 9800001 9910000  
AP008207\_099 9900001 10010000  
AP008207\_100 10000001 10110000  
AP008207\_101 10100001 10210000  
AP008207\_102 10200001 10310000  
AP008207\_103 10300001 10410000  
AP008207\_104 10400001 10510000  
AP008207\_105 10500001 10610000  
AP008207\_106 10600001 10710000  
AP008207\_107 10700001 10810000  
AP008207\_108 10800001 10910000  
AP008207\_109 10900001 11010000  
AP008207\_110 11000001 11110000  
AP008207\_111 11100001 11210000  
AP008207\_112 11200001 11310000  
AP008207\_113 11300001 11410000  
AP008207\_114 11400001 11510000  
AP008207\_115 11500001 11610000  
AP008207\_116 11600001 11710000  
AP008207\_117 11700001 11810000  
AP008207\_118 11800001 11910000  
AP008207\_119 11900001 12010000  
AP008207\_120 12000001 12110000  
AP008207\_121 12100001 12210000  
AP008207\_122 12200001 12310000  
AP008207\_123 12300001 12410000

AP008207\_124 12400001 12510000  
AP008207\_125 12500001 12610000  
AP008207\_126 12600001 12710000  
AP008207\_127 12700001 12810000  
AP008207\_128 12800001 12910000  
AP008207\_129 12900001 13010000  
AP008207\_130 13000001 13110000  
AP008207\_131 13100001 13210000  
AP008207\_132 13200001 13310000  
AP008207\_133 13300001 13410000  
AP008207\_134 13400001 13510000  
AP008207\_135 13500001 13610000  
AP008207\_136 13600001 13710000  
AP008207\_137 13700001 13810000  
AP008207\_138 13800001 13910000  
AP008207\_139 13900001 14010000  
AP008207\_140 14000001 14110000  
AP008207\_141 14100001 14210000  
AP008207\_142 14200001 14310000  
AP008207\_143 14300001 14410000  
AP008207\_144 14400001 14510000  
AP008207\_145 14500001 14610000  
AP008207\_146 14600001 14710000  
AP008207\_147 14700001 14810000  
AP008207\_148 14800001 14910000  
AP008207\_149 14900001 15010000  
AP008207\_150 15000001 15110000  
AP008207\_151 15100001 15210000  
AP008207\_152 15200001 15310000  
AP008207\_153 15300001 15410000  
AP008207\_154 15400001 15510000  
AP008207\_155 15500001 15610000  
AP008207\_156 15600001 15710000  
AP008207\_157 15700001 15810000  
AP008207\_158 15800001 15910000  
AP008207\_159 15900001 16010000  
AP008207\_160 16000001 16110000  
AP008207\_161 16100001 16210000  
AP008207\_162 16200001 16310000  
AP008207\_163 16300001 16410000  
AP008207\_164 16400001 16510000  
AP008207\_165 16500001 16610000  
AP008207\_166 16600001 16710000  
AP008207\_167 16700001 16810000  
AP008207\_168 16800001 16910000  
AP008207\_169 16900001 17010000  
AP008207\_170 17000001 17110000  
AP008207\_171 17100001 17210000  
AP008207\_172 17200001 17310000  
AP008207\_173 17300001 17410000  
AP008207\_174 17400001 17510000  
AP008207\_175 17500001 17610000  
AP008207\_176 17600001 17710000  
AP008207\_177 17700001 17810000  
AP008207\_178 17800001 17910000  
AP008207\_179 17900001 18010000  
AP008207\_180 18000001 18110000  
AP008207\_181 18100001 18210000  
AP008207\_182 18200001 18310000  
AP008207\_183 18300001 18410000  
AP008207\_184 18400001 18510000  
AP008207\_185 18500001 18610000  
AP008207\_186 18600001 18710000  
AP008207\_187 18700001 18810000  
AP008207\_188 18800001 18910000  
AP008207\_189 18900001 19010000  
AP008207\_190 19000001 19110000  
AP008207\_191 19100001 19210000  
AP008207\_192 19200001 19310000  
AP008207\_193 19300001 19410000  
AP008207\_194 19400001 19510000  
AP008207\_195 19500001 19610000  
AP008207\_196 19600001 19710000





ATGOMKPPSSVHEVOKLAGOIASLSRFLSKAERASTLVAVSALVOETEGOKPYVYV  
SKLQCAKTRVYIEKRLAYALWALAKIKGYPQAHAVIAPSOYPLGECILRDKKQRL  
SKRAIAVLPDPDLHFVAFRAIKSOYLDLFAEWTSVLAAPDEPEOSWMLCSDSMSHK  
GAGIAAVLPSPNVPPIRYAARLOFTTNNAEYEAFLGIRAKALGVRLILRIDSK  
LVAGHVEKSEKAEKMGKRYLEAVRMEKCTGITEVHLPRDNEADALASTACGA  
PHSPGJFEFKVIAHSPMDSSEVMAIDQELGDPDMRPYKXLETMLVDEAKA  
KRLQATKATKQVNSGOLYRSGLQPLRLCISFVGEKEMKEIHQGLGHAQVARTYA  
KVRQGVWMPYIVLKVCVEQIKKCESCORHRSQAPQOYLOPLAIPWPAHMQPDIIG  
PPFVANKYKFAIVAVEYRSRWIEAPLCAITSAVOMFYKNIVCRFQVPEKITDN  
GKQFSDDKREKCEGLNLETRFASVALLOSNGAEKTNGLIEALKRLLEGATKQWP  
EELL5VLWLRBPPOPSLAMPALGANSFPMWPSGGEGREVSLEJLEGARYVAL  
EHLKATYSTATSTAYKKNKVPTELMPLGLVLRKKANVAVAKFESKWEFPYLLIKKSR  
EHLATLREGEEDHSMNA5LKRFPYV"

complement (19612..20131)  
/note="5' LTR"  
/protein\_id="BAB39951.1"  
/db\_xref="GI:13486716"

complement (122358..23037,23572..23841,24555..24948)  
/gene="OSJNBa0004B13.5"  
complement (join(122358..23037,23572..23841,24555..24948))  
/gene="OSJNBa0004B13.5"  
/note="contains ESTs AU032146(R3628), AU069558(C62869)"  
/codon\_start=1  
/product="putative tubulin beta-4 chain"  
/protein\_id="BAB39951.1"  
/translation="MRKILHIOGGCGNOIGSKFMEVCDHGIDPTGRVYGTSDLOL  
ERNVYNNASCGRPVPRVAMLDLBEGTMDSVRTGCGTFRPDNRVFGSGAGNMA  
KGYTIGALIDSLVDVVRKEANCDLQFQVCHSLGGGTGSGKTLTISKIRETYP  
DRMLTFVSVPSPKVDVVEPNATLSVHOLENDEKVCNLDNEALYDICRTKLT  
TSFGDLNLHLSATMSGVTCCLFPGQLSDSLKLVNLIPFRLTFVAVGAPLTSR  
GSOQYALVPELTQOMDAKNMCAADPRHGLYIASAMFRKMTKEDCOMINVQ  
NKKSSYFVEMIPNNVKSVCIDIPRGLSMATFSTGSIQEMFRFVSQFLMPFRK  
APLHWYTGSGDMETFEASNNNDVSEYQYQDPTAEBEGIEEBEEQVDE"

join(29477..29557,30999..31115,31504..31614)  
/gene="OSJNBa0004B13.6"  
join(29477..29557,30999..31115,31504..31614)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB39952.1"  
/db\_xref="GI:13486717"

/translation="MDCODLVOKIIGVASTLLANSGLIGHVEGRPOKODLIVSIF  
TSFVQLLLIADHSYGLSHRLSHNNILAMGNGTAVGAMRSRTDRTVTFLLPLADV  
"

complement (join(37256..38344,38502..38665,41064..41640))  
/gene="OSJNBa0004B13.7"  
complement (join(37256..38344,38502..38665,41064..41640))  
/gene="OSJNBa0004B13.7"  
/note="hypothetical protein  
similar to Arabidopsis thaliana chromosome 5, MC9.22"  
/codon\_start=1  
/protein\_id="BAB39953.1"  
/db\_xref="GI:13486718"

/translation="MASPRRSLPLPLLLVFPVSLFVLLLHRRSSIPAEELSGPG  
PPRRSLIKVLYADRPGLRCLSLAADYAGRVALLHVLVDPERNASLDARE  
LIAADALRMPGEEKRVYRANAGLOAQOIEAMWPGSDDEFVVEDLEVTPLYR  
FLKRLVMAVYDRENSPYVFGASLQRPFRVAGKKNKQLDSETHLTYXQVATMGQ  
LLEPKWKEFRMLYDTHSKSGIKFTILEGKKTGWTQKWERLWTFPIFLFVSCGFN  
FTYNFLKERALSVRHDAVNGVRSVGPSTLLDGNLDLPMELQPLMKWYPCF  
AEVLPGRVIRKSELSGVLENNVLLISYLSLEQRIANLLCYLEKSMRYNIF  
LVNDTEFLDLARHGHPVIDAISLQISKMSIYDDPKKEIVAVAYIYKCLDGY  
NLWVNLNGNTISLGSKLNPDSQSVDFEASVDLMEIRGSSKKTWNLDTLRMADG  
MMSKSGFSSSLHKKRVVLTGVLGNNAVRGLKDEITMAVELOPNTSNLSBEGH  
CYLVFWSHMTSDSVQLENKGLMLTIDSSOSAVVCGQKK"

join(42066..42180,43982..44125,45133..45258,45484..45602,  
45939..46208)  
/gene="OSJNBa0004B13.8"  
join(42066..42180,43982..44125,45133..45258,45484..45602,  
45939..46208)  
/gene="OSJNBa0004B13.8"  
/note="contains ESTs AU062927(C51629), AU030693(B60120)  
similar to Arabidopsis thaliana chromosome 1, FlAP1.8  
unknown protein"  
/codon\_start=1

/protein\_id="BAB39954.1"  
/db\_xref="GI:13486719"  
/translation="MAGSKAFYANGFWIRETGALDLGCRLOGNTYFHEQLSHRT  
LMTNIDKPHHRAVAPASLSLIGDIVGGASITWYCVLRGDNANNQISGNIQD  
NSLVHANSNLSGKFPPTIIDNVTVGHSAVLQGTCTVEDEFVGGATLLDGVVEKH

Query Match 11.6%; Score 57; DB 15; Length 142268;  
Best Local Similarity 55.2%; Pared No. 0.03;  
Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

1 CGCGCGCGCGCGGAGATGCCAGATGCCAAGAGCCGAGTTTGAGAGCTGTGGCAGAAAT 60  
Db 104620 CCACCGAGCTTGCGCAGACCTTGCTTGAGAGTGTCAGCGGACATGTGTGAGCGGAGG 104561

Qy 61 TCCCGGAGAGAGAGCCACGCTGTCTCTGTGTGGAAGCCGAGGCTGCTGTGTGATGAGA 120  
Db 104560 ACCCGGAGAGATGTGTGACGACGAGAGTGCGCGAGACCTGAGAGATGTGTGCGAGAG 104501

Qy 121 ACTCTTTCGCGAGGAGTGCTGCGGTGTGACGAGAGAGCTCAGAGGCTGCGACAGTCTG 180  
Db 104500 ACGTGTCCGAGGAGACCGAGAGATGTGTGACGAGAGATGCGGAGCCGAGAGATGG 104441

Qy 181 GCGCGGCGCTTGAGGCTGCTGG 201  
Db 104440 TGCACGAGATGCGCTTGTG 104420

RESULT 13  
AP003631/c  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
PAC clone: p0581P09.  
ACCESSION  
AP003631 BA000010  
VERSION  
AP003631.3 GI:15408812  
KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzae; Oryza.

REFERENCE  
AUTHORS  
1 Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,  
Karayose, Y., Wu, J., Niiimura, Y., Cheng, Z., Nagamura, Y.,  
Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Ariawa, K.,  
Chidem, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arida, K.,  
Hamada, M., Harada, C., Hishida, S., Honda, M., Ichikawa, Y.,  
Idomura, A., Iijima, M., Ikeda, M., Ikemoto, S., Ito, T., Ito, Y.,  
Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, M., Katagiri, S.,  
Kikuta, A., Kobayashi, N., Kono, I., Machida, K., Maehara, T.,  
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,  
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,  
Ohta, I., Ono, N., Saiji, S., Sakai, K., Shibata, M., Shimokawa, T.,  
Shimura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,  
Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,  
Zhong, H., Iwama, H., Endo, T., Ito, H., Hann, J. H., Kim, H. I., Eun, M. Y.,  
Yano, M., Jiang, J. and Gojobori, T.  
The genome sequence and structure of rice chromosome 1  
Nature 420 (6913), 312-316 (2002)  
12447438  
2 (bases 1 to 154248)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (17-MAY-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki, 305-8602, Japan  
(E-mail: tsasaki@ias.affrc.go.jp, URL: http://gsp.dna.affrc.go.jp/  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
On Aug 31, 2001 this sequence version replaced gi:14861131.  
Genes were predicted from the integrated results of the following:  
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH  
(http://www.softberry.com/), Genemark.hmm  
(http://opal.biology.gatech.edu/Genemark/), Glimmer  
(http://www.cigr.org/tcd/glimmer/glimr\_form.html), RiceHMM

(http://rtp.dna.affrc.go.jp/RiceHm/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://glabin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (http://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from 5' to 3' of the PAC clone. This sequence of P0581F09 clone has an overlap with P0498A12 (DDBJ: AF003020) clone at 3' end. Detailed information on overlap and assembly qualify together with annotation of this entry is available at <http://rtp.dna.affrc.go.jp/GenomeSeq.html>.

## FEATURES

## source

Location/Qualifiers

1. 154248

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/culivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="1"

/clone="P0581F09"

2373..3422

/gene="P0581F09.1"

2373..3422

/gene="P0581F09.1"

/note="probably inactive due to 5'exon missing in CDS  
probably inactive due to including stop codon(6) in CDS  
pseudogene, Retrovirus-related Pol polypeptide from  
transposon TNT 1-94"

/pseudo

complement(3564..4586)

/gene="P0581F09.2"

complement(join(3564..3818,4200..&gt;4586))

/gene="P0581F09.2"

/note="start and end point are not identified"

complement(join(3564..3818,4200..4586))

/gene="P0581F09.2"

/note="predicted by GeneMark.hmm etc."

/codon\_start=1

/product="hypothetical protein"

/protein\_id="BAD82601.1"

/db\_xref="GI:56785019"

/translation="MGVLPDRDEACRPLQMAKPYLMGKRGARDDIDVHYIVSIK  
HLSPHVAIVLHNDSDHPKQYFHRQISEKRWKATISGHPAPAIINLSSKYLIN  
KINQINDPLKKNHDANDPTIRWEKEREKRELDGDELDLVGTASPRSSITTKEN  
NSGPTTGTGRISTEGTSSNLASTWATORSILVABASSILPATGLSTLAAGLY"

complement(5017..5442)

/gene="P0581F09.3"

complement(5017..5442)

/gene="P0581F09.3"

/note="start and end point are not identified"

complement(5017..5442)

/gene="P0581F09.3"

/note="predicted by GeneMark.hmm etc."

/codon\_start=1

/product="hypothetical protein"

/protein\_id="BAD82602.1"

/db\_xref="GI:56785020"

/translation="MLGMRPCSGGLPPIHRORLLESAGCGGAGKAGADPAALVYGR  
GSGACLPFGDGDGAREGRCGSLPIPRRMFSSAAGGGGTGGADPAALVYGR  
AGSGACLPFGGGGSGRRSGMGWIRRRSSLDSDSD"

## gene

## mRNA

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

CDS	join(24883..24998,25151..25535)
	/gene="P0581F09.8"
	/note="Predicted by GeneMark.hmm etc."
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="BAD82605.1"
	/db_xref="GI:56785023"
	/translation="MATATYADSELYVTVPKDAIPDDDDGGDAASRGMRLRSLVY
	GDPLEPADGEGRRGRDAGBERGMDASACGERRGRDSAGEGRRGRMDAGVGE
	SASDEPHRDVMSGAPARMTCSRRYRSVPAASVPHPHAMREKRRERREGR
Query Match	11.6%; Score 57; DB 15; Length 154248;
Best Local Similarity	55.2%; Pred. No. 0.03; Mismatches 90; Indels 0; Gaps 0;
Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;	
Qy	1 CGGCGCGCGCGCGGAGTCCGAGTCCGAGGCGGAGCTTTGAGAGCGTGTGGCAGAA 60
Db	127490 CGACGAGAGCTTGGCGAGACCTTGTCTGGAGGTGCGGCGGACATGTGTGAGCGGAGG 127431
Qy	61 TCCCGGAGAGAGGCGCCAGCTCTCCCTGTGTGAGAGCGGCGGCTGTGTGTATGAGA 120
Db	127430 AGCGGAGAGAGTGTGACAGAGAGTGGCGGAGAGCTGTGAGAGATGTGTCAAGAGG 127371
Qy	121 AGCTTCTCCGAGAGGTGTCTCCGTGTGTGAGAGAGCTCAGAGAGCTTGGCAGAGTCGT 180
Db	127370 ACCTGCGGAGAGAGACCGGAGAGATGTGTGAGAGAGAGTGGCGGAGCGGAGAGATGG 127311
Qy	181 GCGCGGCGCTTGAAGCTGCTGG 201
Db	127310 TGCACGAGATGCCGCTTGTGG 127290
RESULT 14	
AP003020/c	159749 bp DNA linear PLN 22-DEC-2004
LOCUS	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
DEFINITION	PAC clone: P0498A12.
ACCESSION	AP003020 BA000010
VERSION	AP003020.2 GI:13486738
KEYWORDS	
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group)
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
	Eurharitideae; Oryzaceae; Oryza.
REFERENCE	
AUTHORS	1 Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
	Katayose, Y., Wu, J., Mitamura, Y., Cheng, Z., Nagamura, Y.,
	Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
	Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
	Hamada, M., Harada, C., Hijiishi, S., Honda, M., Ichikawa, Y.,
	Idonuma, A., Iijima, M., Ikeda, M., Ikeda, M., Ito, S., Ito, T., Ito, Y.,
	Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, M., Katagiri, S.,
	Kikuta, A., Kobayashi, N., Kono, I., Machida, K., Maehara, T.,
	Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
	Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
	Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,
	Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Teuji, K., Waki, K.,
	Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,
	Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
	Yano, M., Jiang, J. and Gojobori, T.
TITLE	The genome sequence and structure of rice chromosome 1
JOURNAL	Nature 420 (6913), 312-316 (2002)
PUBMED	12447438
REFERENCE	2 (bases 1 to 159749)
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE	Direct Submission
JOURNAL	Submitted (06-DEC-2000) Takuji Sasaki, National Institute of
	Agricobiological Sciences, Rice Genome Research Program, Kannondai
	2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
	(E-mail: tsasak@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
	Tel: 81-298-38-7441, Fax: 81-298-38-7468)
COMMENT	On Mar 28, 2001 this sequence version replaced gi:11602827.

FEATURES	
source	1..159749
	/organism="Oryza sativa (japonica cultivar-group)"
	/mol_type="genomic DNA"
	/cultivar="Nipponbare"
	/db_xref="taxon:39947"
	/chromosome="1"
gene	1..333
	/gene="P0498A12.1"
	<1..>333
	/note="start and end point are not identified"
mRNA	1..333
	/gene="P0498A12.1"
	/note="contains full-length cDNA (s): AK119993"
CDS	1..333
	/gene="P0498A12.1"
	/note="contains full-length cDNA (s): AK119993"
	/codon_start=1
	/product="unknown protein"
	/protein_id="BAD81552.1"
	/db_xref="GI:56784167"
	/translation="MEMGYSSEERKPEKDTSRSAVSFSSSSSAGTRPSCWRS
	VTTTARRRCCGDEHDTAARASVLAHDHHLGIGIKVDSAGAVRTLVERNDPYC
	DECNTHGR"
	complement(join(1116..2215,3205..3549))
	/gene="P0498A12.2-1"
	complement(join(1116..2215,3205..3549))
	/gene="P0498A12.2-1"
	complement(join(1116..2215,3205..3549))
	/note="supported by full-length cDNA (s): AK064158"
	complement(join(1424..2215,3205..3465))
	/gene="P0498A12.2-1"
	/note="contains EST (s): AU166951 (CG0192), AU166950 (CG0192)
	contains full-length cDNA (s): AK064158
	similar to Arabidopsis thaliana chromosome 3, AC3g17615"
	/codon_start=1
	/product="unknown protein"
	/protein_id="BAB39974.1"
	/db_xref="GI:13486740"
	/translation="MGAGMSSGRRPSFYGGGGGVGRPRGMLPLLAQLVLEYGAG
	ASRPVTAALAAANLTYLRPGSLDALPLINRVAFNPHLIIHYCDLTRFLSAYHL

Genes were predicted from the integrated results of the following:  
GENSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), FGENESH  
(<http://www.softberry.com/>), GeneMark.hmm  
(<http://opal.biology.gatech.edu/GeneMark/>), GlimmerM  
([http://www.tigr.org/tdb/glimmerm/glmr\\_form.html](http://www.tigr.org/tdb/glimmerm/glmr_form.html)), RiceHMM  
(<http://rgp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor  
(<http://biointformatics.iastate.edu/cgi-bin/isp.cgi>), sstm4  
(<http://globin.cse.psu.edu/html/doc/sim4.html>), gap2  
(<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The  
genomic sequence was searched against NCBI Nonredundant Protein  
database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA  
sequence database at RGP or DBJ. Protein homologs of the coding  
regions were searched against NCBI Nonredundant Protein database  
with BLASTP. ESTs represent the identified cDNA sequences using  
BLASTN with the corresponding DBJ accession no. and RGP clone ID.  
Full-length cDNAs represent the identified cDNA sequences using  
BLASTN with the corresponding DBJ accession no.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with full-length cDNA or  
EST homology (covering almost the entire length of partial  
sequence) is classified as an 'unknown' protein. A gene predicted  
by two or more gene prediction programs is classified as a  
'hypothetical' protein according to IRGSP standard. A gene  
predicted by a single gene prediction program is also classified as a  
probable 'hypothetical' protein and is included as a  
miscellaneous feature of the sequence.  
The orientation of the sequence is from SP6 to T7 of the PAC clone.  
This sequence of P0498A12 clone has an overlap with P0581F09 (DBJ:  
AP003631) clone at 5' end and with P0511C01 (DBJ: AP002070) clone  
at 3' end. Detailed information on overlap and assembly quality  
together with annotation of this entry is available at  
<http://rgp.dna.affrc.go.jp/Genomeseq.html>.

SETHPEPNSSLLMKGCIQLETSKGSVFASMAALIGMSQGITLLLSKLLIFGNDEA  
YVDQYAVGSGVLFQMKVVLNMSDYLHGVLPKAYAAAEILLIGAFIPGSLI  
GHLGGLAGIAYLMLKRSFSDPDLISLISGAKARMPVGFQKLFSGRGYTPS  
BRVVGSGASRENGRIGWCSACTYDNPSTDICEMCSSAREDHAFSHRQHLQAGNGE  
PSVEEIRRRRLERFSR"  
/gene="P0498A12.2-2"  
/complement(42582..>3556)  
/note="supported by full-length cDNA(s) : AKI07821"  
/complement(2582..3556)  
/gene="P0498A12.2-2"  
/note="contains full-length cDNA(s) : AKI07821  
non-coding transcript  
probably inactive due to including stop codon(s) in CDS"  
complement(6026..6223)  
/gene="P0498A12.3"  
/complement(6026..6223)  
/gene="P0498A12.3"  
/note="hypothetical ORF  
predicted by GlimmerM  
this category is not included in IRGSP standard"  
761..9681  
/gene="P0498A12.4"  
761..9681  
/gene="P0498A12.4"  
/note="probably inactive due to including stop codon(s) in  
CDS  
pseudogene, cinnamoyl-CoA reductase"  
pseudo  
12167..15670  
/gene="P0498A12.5"  
join(12167..12381,13185..13339,13451..13633,14762..15114,  
15350..15670)  
/gene="P0498A12.5"  
/note="supported by full-length cDNA(s) : AKI07149"  
join(12258..12381,13185..13339,13451..13633,14762..15114,  
15350..15521)  
/gene="P0498A12.5"  
/note="contains EST(s) : AU095269(EB1006)  
contains full-length cDNA(s) : AKI07149"  
/codon\_start=1  
/product="putative cinnamoyl-CoA reductase"  
/protein\_id="BAB39976.1"  
/db\_xref="GI:13486742"  
/translation="MYTGRSEQWCVTGAAGTGSMTLYKELHARGYFVRGAREPADI  
KNALHVLDAERGLSLRADVDRLRAAFALCDGVFHYASVNSDEILPAIEIG  
TKNVLNAADMGVRRVVTSSYGAHNPFRSDQIVDESCWDLFCQOTQNNYCTA  
KMLAERTAMEASRKGYNLLVVPVAVGEMLOPTLASVHRVATVWRGKSAFPAV  
AAVVDVDAARALVYEHDPARGRYLCISVLRSEFVRLRELPOYITSRCKDN  
SKMVKPKRSVQLETLGMQFPLKESIVRTVISLQDKHLPAISRSLM"  
19216..19533  
/gene="P0498A12.6"  
19216..19533  
/gene="P0498A12.6"  
/note="start and end point are not identified"  
19216..19533  
/gene="P0498A12.6"  
/note="predicted by GlimmerM etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD81554.1"  
/db\_xref="GI:56784168"  
/translation="MLGKHPSSPDYIKTKTLTPTRPRSKOPPTAPAPVILP  
LHPASSKMTAKAKKPKRGRHVHCGSMETPQWRBGMGRDIEGRVAEDIALVY  
IDL"  
19808..22072  
/gene="P0498A12.7"  
join(<19808..19955,21771..21820,21986..22072)  
/gene="P0498A12.7"  
/note="start and end point are not identified"  
join(19808..19955,21771..21820,21986..22072)  
/gene="P0498A12.7"

/note="predicted by Genemark.hmm etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD81554.1"  
/db\_xref="GI:56784169"  
/translation="MGAVODGSSSDGRVNGENHARSLSLDGGGPRKVGREBEAT  
ALSVGRIGSGGINKATTTKIGAFPCGLMSVNAERIGVILIGYSSLSKG"  
22938..23593  
/gene="P0498A12.8"  
join(<22938..23027,23399..23486,23586..>23593)  
/gene="P0498A12.8"  
/note="start and end point are not identified"  
join(22938..23027,23399..23486,23586..23593)  
/gene="P0498A12.8"  
/note="predicted by GENSCAN etc."  
Query Match 11.6%; Score 57; DB 15; Length 159749;  
Best Local Similarity 55.2%; Pred. No. 0.03;  
Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
CDS  
gene  
mRNA  
1 CGGCGCGCCCGGGGAGTCCGAGTCCCAAGAGCCGAGTTTGAAGAGCTGTGTGCAGAAAT 60  
4535 CGACGAGCTTGGCGAGACTTGTCTTGGAGGTGCGAGCGGACATGTGAGAGCGAGG 45276  
QY 61 TCCCGAGAAAGAGCGCCAGCTGTCTCCGTGTGGAAGCGAGGGCTGTGTATGAGAA 120  
Db 45275 AGCCGAGAGATGTGTGACAGCAGAGCTGTGCGAGAGCTGTGAGAGATGTGCAGAGG 45216  
QY 121 AGCTTCTCCGAGGAGTGTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
Db 45215 ACCTGCCGAGAGAACCGAGAGAGATGTGTGAGAGAGATGTGTGAGAGAGATGTG 45156  
QY 181 GCGCGGCTTGAAGCTGTGTG 201  
Db 45155 TGCACAGAGATGCCGCTTGTG 45135  
RESULT 15  
AY571308 2279 bp mRNA linear INV 10-JUN-2004  
LOCUS Kukulcania hibernalis major ampullate spidroin 1 mRNA, partial cds.  
DEFINITION AY571308  
ACCESSION AY571308  
VERSION AY571308.1 GI:47007922  
KEYWORDS  
SOURCE Kukulcania hibernalis  
ORGANISM Kukulcania hibernalis  
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
Araneomorphae; Haplogynae; Filistatidae; Kukulcania.  
REFERENCE 1 (bases 1 to 2279)  
Tian,M., Liu,C. and Lewis,R.  
AUTHORS Analysis of major ampullate silk cDNAs from two non-ord-weaving  
spiders  
TITLE Biomacromolecules 5 (3), 657-660 (2004)  
JOURNAL 15132643  
PUBMED 2 (bases 1 to 2279)  
REFERENCE Liu,C., Tian,M. and Lewis,R.  
AUTHORS Direct Submission  
TITLE Submitted (05-MAR-2004) Molecular Biology, University of Wyoming,  
JOURNAL 1000 E. University Avenue, Laramie, WY 82071-3944, USA  
FEATURES  
source  
1..2279  
/organism="Kukulcania hibernalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:268415"  
<1..2279  
/codon\_start=1  
/product="major ampullate spidroin 1"  
/protein\_id="AAI08433.1"  
/db\_xref="GI:47007923"  
/translation="GSSGDGASAAASAAAAGAGSGRRGPGARSRGAGAGAGAGCG  
VGVGSGSAGAGAGAGAGAGCGEGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  
SASASAAVVFESGAG  
GRGRGGEAHSASASAAVVFESAGPEEAGSSGDGASAAASAAAAGAGSGRRGPGC



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 03:01:41 ; Search time 466 Seconds  
(without alignments)  
7050.844 Million cell updates/sec

Title: US-10-071-510A-16  
Perfect score: 493  
Sequence: 1 cggcgcggccggggagatgcc.....cctgtctcatttgagcctgc 493

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues  
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :  
1: Genesegq1980s:\*  
2: Genesegq1990s:\*  
3: Genesegq2000s:\*  
4: Genesegq2001as:\*  
5: Genesegq2001bs:\*  
6: Genesegq2002as:\*  
7: Genesegq2002bs:\*  
8: Genesegq2003as:\*  
9: Genesegq2003bs:\*  
10: Genesegq2003cs:\*  
11: Genesegq2003ds:\*  
12: Genesegq2004as:\*  
13: Genesegq2004bs:\*  
14: Genesegq2005s:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	100.0	493	10	ADH69132 Human tum
2	347.6	70.1	4083	5	AAS83954 DNA encod
3	345.4	70.1	3003	12	ADG67408 Novel hum
4	176.4	35.8	532	12	ACH77474 Human gen
5	174	35.3	174	12	ACH91174 Human gen
6	56.8	11.5	2282	8	AA151693 Kukulicani
7	51.6	10.5	3711	13	ADU01598 Novel hum
8	51.4	10.4	1505	2	AA055750 Genomic c
9	50.2	10.2	15231	6	ABK63598 Rat seque
10	50.2	10.2	15231	10	ADBS59205 Toxicity-
11	50.2	10.2	15231	10	ADBS59206 Primary r
12	50.2	10.2	15231	10	ADBS59207 Toxicity
13	50.2	10.2	15231	10	ADBS59208 Renal tox
14	49.8	10.1	768	6	AB213007 Arabidops
15	49.8	10.1	2328	6	ADU01973 Novel hum
16	49.8	10.1	16442	2	AAK72613 Human imm
17	49.4	10.0	8298	4	AAK72613 Human imm
18	49	9.9	8973	12	ADP28653 Human sec
19	48.6	9.9	114793	4	AA088215 Human gen

20	48.2	9.8	628	12	ACH89664 Human gen
21	47.8	9.7	799	2	AAV55831 Nucleotid
22	47.8	9.7	913	2	AAI07189 Cotton fl
23	47.8	9.7	913	2	AAI13034 Cotton fl
24	47.8	9.7	913	2	AAI30253 Cotton fl
25	47.8	9.7	913	2	AAI30253 Cotton fl
26	47.8	9.7	913	2	AAI62610 Cotton fl
27	47.8	9.7	913	2	AAI70041 Cotton fl
28	47.8	9.7	913	2	AAI70041 Cotton fl
29	47.8	9.7	913	2	AAI70041 Cotton fl
30	47.8	9.7	913	2	AAI70041 Cotton fl
31	47.8	9.7	913	2	AAI70041 Cotton fl
32	47.8	9.7	913	2	AAI70041 Cotton fl
33	47.8	9.7	913	2	AAI70041 Cotton fl
34	47.8	9.7	913	2	AAI70041 Cotton fl
35	47.8	9.7	913	2	AAI70041 Cotton fl
36	47.8	9.7	913	2	AAI70041 Cotton fl
37	47.8	9.7	913	2	AAI70041 Cotton fl
38	47.8	9.7	913	2	AAI70041 Cotton fl
39	47.8	9.7	913	2	AAI70041 Cotton fl
40	47.8	9.7	913	2	AAI70041 Cotton fl
41	47.8	9.7	913	2	AAI70041 Cotton fl
42	47.8	9.7	913	2	AAI70041 Cotton fl
43	47.8	9.7	913	2	AAI70041 Cotton fl
44	47.8	9.7	913	2	AAI70041 Cotton fl
45	47.8	9.7	913	2	AAI70041 Cotton fl

ALIGNMENTS

RESULT 1  
ADH69132 standard; DNA; 493 BP.  
ID ADH69132;  
AC ADH69132;  
AC 25-MAR-2004 (first entry)  
DT Human tumour resistance/sensitivity marker DNA #16.  
DE human; ds; tumour; tumour growth; cancer; resistance; sensitivity.  
KW Homo sapiens.  
XX US2003143552-A1.  
XX 31-JUL-2003.  
XX 08-FEB-2002; 2002US-00071510.  
XX 08-FEB-2001; 2001US-0267276P.  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Clark E, Grenfell-Lee T, Lu K, Hartmann L, Brown JL;  
XX WPI; 2003-829783/77.  
XX Determining whether agent can/cannot be used to reduce growth of tumor  
XX involves exposing obtained tumor cells to test agent, determining  
XX expression level of markers by tumor cells.  
XX Disclosure; SEQ ID NO 16; 41pp; English.  
XX The invention relates to a method of determining whether an agent  
XX can/cannot be used to reduce the growth of tumour, involves obtaining a  
XX sample of tumour cells, exposing them to one or more test agents,  
XX determining the level of expression of one or more markers by the cells  
XX exposed to the agent and by the cells not exposed to the agent and  
XX identifying the agent as appropriate/inappropriate to reduce the growth  
XX of tumour. The method is useful for determining whether an agent  
XX can/cannot be used to reduce the growth of tumour. The method is useful

CC for determining whether an anti-cancer agent should in treatment of a  
CC cancer patient should be continued/discontinued. The method is useful for  
CC reducing growth rate of cancer in a patient. The present sequence  
CC represents a human tumour resistance/sensitivity DNA marker.  
XX

Sequence 493 BP, 99 A; 124 C; 174 G; 96 T; 0 U; 0 Other;

Query Match 100.0%; Score 493; DB 10; Length 493;  
Best Local Similarity 100.0%; Pred. No. 7,2e-114;  
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CGGCGCGCGCGGAGATGCCAGTCCCAAGAGCGGAGTTTGAAGAGCTGTGGCAGAAT 60
DB 1 CGGCGCGCGCGGAGATGCCAGTCCCAAGAGCGGAGTTTGAAGAGCTGTGGCAGAAT 60
QY 61 TCCCGGAGAAAGAGCCCACTGTCTCTGTGGAAGCCAGAGGCTGTGTATGAGAGA 120
DB 61 TCCCGGAGAAAGAGCCCACTGTCTCTGTGGAAGCCAGAGGCTGTGTATGAGAGA 120
QY 121 AGTCTTCTCGGAGGGGTGCTGCGTGTGAGAGAGGCTCAGGAGCTGGCAGAGTCT 180
DB 121 AGTCTTCTCGGAGGGGTGCTGCGTGTGAGAGAGGCTCAGGAGCTGGCAGAGTCT 180
QY 181 GCGCGGCTTGAAGGCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGC 240
DB 181 GCGCGGCTTGAAGGCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGC 240
QY 241 AGAGGATGGAAGTGTGATTCGGGGAGAAATGTTTTCACCAACACATCCCAAAGTCAG 300
DB 241 AGAGGATGGAAGTGTGATTCGGGGAGAAATGTTTTCACCAACACATCCCAAAGTCAG 300
QY 301 GATTTCATCAATCCATGATTCCTATTCACAGGCAATGTCGAGCGGTGAGTCTGCTA 360
DB 301 GATTTCATCAATCCATGATTCCTATTCACAGGCAATGTCGAGCGGTGAGTCTGCTA 360
QY 361 GCAGGCGCTGTGGAAGAGGCGCAGGCGCCAGGTCAAGAGTGGAGTCTCCAGCA 420
DB 361 GCAGGCGCTGTGGAAGAGGCGCAGGCGCCAGGTCAAGAGTGGAGTCTCCAGCA 420
QY 421 CAGGCGCTTCTGTGTCTGCGGCAACATGTCTGTCTGAGACTTGGCCAGTCTGTCT 480
DB 421 CAGGCGCTTCTGTGTCTGCGGCAACATGTCTGTCTGAGACTTGGCCAGTCTGTCT 480
QY 481 CATTGAGCCTGC 493
DB 481 CATTGAGCCTGC 493
```

RESULT 2  
AAS83954  
ID AAS83954 standard; cDNA; 4083 BP.  
XX  
AC AAS83954;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #19758.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
XX  
PR 23-AUG-2000; 2000US-00649167.  
XX  
F PA (HYSE-) HYSEQ INC.

XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI, 2001-639362/73.  
XX  
DR P-PSDB; ABG19767.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnosis, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

Claim 1; SEQ ID NO 19758; 103bp; English.

XX  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostic, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

Sequence 4083 BP, 923 A; 1177 C; 1228 G; 755 T; 0 U; 0 Other;

Query Match 70.5%; Score 347.6; DB 5; Length 4083;  
Best Local Similarity 98.9%; Pred. No. 4,2e-77;  
Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 2 GCGCGGCGCGGAGATGCCAGTCCCAAGAGCGGAGTTTGAAGAGCTGTGGCAGAATT 61
DB 2796 GCGCGGCGCGGAGATGCCAGTCCCAAGAGCGGAGTTTGAAGAGCTGTGGCAGAATT 2855
QY 62 CCGGAGAAAGAGGCCAGTCTCTCTGTGGAAGCGCAGGCTGTGTGATGAGAGA 121
DB 2856 CCGGAGAAAGAGGCCAGTCTCTCTGTGGAAGCGCAGGCTGTGTGATGAGAGA 2915
QY 122 GTCTTCTCCGAGAGGTGCTGCGGTGTGTCAGAGAGAGTCAAGAGCTGGCAGAGTCTG 181
DB 2916 GTCTTCTCCGAGAGGTGCTGCGGTGTGTCAGAGAGAGTCAAGAGCTGGCAGAGTCTG 2975
QY 182 GCGGCGCTTGAAGGCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGCA 241
DB 2976 GCGGCGCTTGAAGGCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGCA 3035
QY 242 GAGATGGAAGTGAATTCGGGGAGAAATGTTTTCACCAACAACATCCCAAAGTCAG 301
DB 3036 GAGATGGAAGTGAATTCGGGGAGAAATGTTTTCACCAACAACATCCCAAAGTCAG 3095
QY 302 ATTTCTCATCAATCCCATGATCCATTCACAGAGATGTGAGCGCGGAGTCT 355
DB 3096 ATTTCTCATCAATCCCATGATCCATTCACAGAGATGTGAGCGCGGAGTCT 3149
```

RESULT 3  
AD067408  
ID AD067408 standard; cDNA; 3003 BP.  
XX  
AC AD067408;  
XX  
DT 07-OCT-2004 (first entry)



XX Novel human cDNA sequence #2381.

DE ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;  
 XX cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;  
 KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;  
 KW cancer.

XX Homo sapiens.

XX EPI440981-A2.

XX 28-JUL-2004.

XX 21-JAN-2004; 2004EP-00001196.

XX 21-JAN-2003; 2003JP-00102206.

XX 09-MAY-2003; 2003JP-00131392.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isegai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;  
 XX Yamamoto J, Isono Y, Nagai K, Irie R;

XX WPI: 2004-535376/52.

XX P-PSDB; ADQ67715.

XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,  
 XX Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

XX Claim 1; SEQ ID NO 4569; 2449pp; English.

XX The invention relates to 2495 novel polynucleotides (I) and their encoded  
 XX polypeptides, sequences hybridizing to these nucleotides, sequences  
 XX encoding partial polypeptides and sequences having 70% or 90% identity to  
 XX the nucleotide and protein sequences. The nucleotides and polypeptides  
 XX are useful as diagnostic markers or therapeutic target for the diseases  
 XX or morbid states. They are also useful for treating osteoporosis,  
 XX neurological diseases, Alzheimer's diseases, Parkinson's diseases,  
 XX dementia and various cancers. This sequence corresponds to a nucleotide  
 XX sequence of the invention.

XX Sequence 3003 BP; 664 A; 817 C; 988 G; 534 T; 0 U; 0 Other;

XX Query Match 70.1%; Score 345.4; DB 12; Length 3003;

XX Best Local Similarity 99.7%; Pred. No. 1.4e-76;

XX Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX 2 GCGCGGCGCGGGGATGCCAGTCCCAAGAGCGCGAGTTTGAAGAGCTGTGTGGCAAAATT 61

XX 1911 GCGCGGCGCGGGGATGCCAGTCCCAAGAGCGCGAGTTTGAAGAGCTGTGTGGCAAAATT 1970

XX 62 CCGCGAGAGAGGCGCCAGCTGTCTCTGTGTGAGAGCGGAGCTGTGTGTGTGTGAGAA 121

XX 1971 CCGCGAGAGAGGCGCCAGCTGTCTCTGTGTGAGAGCGGAGCTGTGTGTGTGTGAGAA 2030

XX 122 GTCTTCTCCGAGGGTGTCTCCGTGTGTGAGAGCGGAGCTGTGTGTGTGTGAGAA 181

XX 2031 GTCTTCTCCGAGGGTGTCTCCGTGTGTGAGAGCGGAGCTGTGTGTGTGTGAGAA 2090

XX 182 GCGGCGCTTGAAGGCTGTGTGAGAGCGGAGCTGTGTGTGTGTGAGAACTGTGCA 241

XX 2091 GCGGCGCTTGAAGGCTGTGTGAGAGCGGAGCTGTGTGTGTGTGAGAACTGTGCA 2150

XX 242 GAGGATGGAAGTGTATTCGGGGAAGAAATGTTTTCACCAACATCTCCCAATTCAG 301

XX 2151 GAGGATGGAAGTGTATTCGGGGAAGAAATGTTTTCACCAACATCTCCCAATTCAG 2210

XX 302 ATTTCATCATCCATCCATGATCTTATCCAGGAGCTGTGTGAGCGG 348

XX 2211 ATTTCATCATCCATCCATGATCTTATCCAGGAGCTGTGTGAGCGG 2257

RESULT 4

ACH77474/c

ACH77474 standard; DNA; 532 BP.

29-JUL-2004 (first entry)

Human genome derived single exon probe #10665.

Human; probe; ss; gene expression; single exon probe; microarray;

alternative splicing event; genomic alteration.

Homo sapiens.

US2003194704-A1.

16-OCT-2003.

03-APR-2002; 2002US-00029386.

03-APR-2002; 2002US-00029386.

(PENN/) PENN S G.

(RANK/) RANK D R.

(HANZ/) HANZEL D K.

Penn SG, Rank DR, Hanzel DK;

WPI: 2004-119264/12.

New human genome-derived single exon nucleic acid probes useful for human

gene expression analysis, for identifying or characterizing alternative

splicing events, for assessing genomic alterations or as tools for

surveying tissues.

Claim 15; SEQ ID NO 10669; 80pp; English.

The invention relates to a nucleic acid probe for measuring human gene  
 expression, comprising any of the 27,400 fully defined nucleotide  
 sequences in the specification, or their complements or fragments, and  
 encoding at least 8 amino acids of any of the 6888 amino acid sequences  
 fully defined in the specification. The probe is a single exon probe that  
 hybridizes under high stringency conditions to a nucleic acid molecule  
 expressed in human cells or tissues. Also included are a spatially-  
 addressable set of single exon nucleic acid probes for measuring human  
 gene expression (comprising a plurality of single exon nucleic acid  
 probes cited above, where each of the plurality of probes is separately  
 and addressably isolatable or amplifiable from the plurality), a single  
 exon microarray for measuring human gene expression, a method of  
 measuring human gene expression, a vector comprising the single exon  
 probe cited above, an ORF-encoded peptide comprising at least 8  
 contiguous amino acids of any of the above-mentioned amino acid  
 sequences (optionally with conservative amino acid substitutions), an  
 isolated antibody that binds specifically to a peptide cited above,  
 methods of selling and/or licensing single exon probes or microarrays to  
 a customer desiring to measure gene expression, a method of providing  
 human gene expression data by subscription, and a computer-readable  
 storage medium which contains a database having a plurality of records  
 (each record including data on the expression of a single exon probe  
 cited above. The probe, methods and apparatus are useful in gene  
 expression analysis. The probes may be used as tools for surveying  
 tissues to detect the presence of expressed messages that contain their  
 specific exon, or in constructing genome-derived single exon microarrays.  
 In addition, the probes are used in identifying and characterizing  
 alternative splicing events, in detecting and characterizing gross  
 alterations in the genomic locus that includes their exon, in assessing  
 smaller genomic alterations, in printing the synthesis of nucleic acids,  
 or in expressing the ORF-encoded peptide. The present sequence is a human  
 single exon probe of the invention. Note: The sequence data for this  
 patent did not form part of the printed specification, but was obtained  
 in electronic format directly from USPTO at  
 seqdata.uspto.gov/sequence.html?docID=20030194704

XX Sequence 532 BP; 113 A; 175 C; 147 G; 97 T; 0 U; 0 Other;  
SQ  
Query Match 35.8%; Score 176.4; DB 12; Length 532;  
Best Local Similarity 99.4%; Pred. No. 2,4e-34;  
Matches 177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 40 TTGAGAGGCTGTGGCAGAAATTCCTCCGAGAGAGGCCCAAGCTGTCCCTGTGGAAGCGC 99  
DB 425 TTTAGAGGCTGTGGCAGAAATTCCTCCGAGAGAGGCCCAAGCTGTCCCTGTGGAAGCGC 366  
QY 100 AGGCGTGTGTGTATGAGAGAAATCTTCTCCGAGAGGCTGTGCTGTGTCAGAGAGAGC 159  
DB 365 AGGCGTGTGTGTATGAGAGAAATCTTCTCCGAGAGGCTGTGCTGTGTCAGAGAGAGC 306  
QY 160 TCAGAGAGCTGTGCAAGTCTGTGGCGGCTTTGAGGCTGTCTGGAAGAAATCTGTCTGAG 217  
DB 305 TCAGAGAGCTGTGCAAGTCTGTGGCGGCTTTGAGGCTGTCTGGAAGAAATCTGTCTGAG 248  
RESULT 5  
ACH91174/c  
ID ACH91174 standard; DNA; 174 BP.  
AC ACH91174;  
XX  
XX 29-JUL-2004 (first entry)  
DT  
XX  
XX Human genome derived single exon probe #24369.  
DE  
XX  
XX Human; probe: ss; gene expression; single exon probe; microarray;  
KM alternative splicing event; genomic alteration.  
XX  
XX Homo sapiens.  
OS  
XX  
XX US2003194704-A1.  
PN  
XX  
XX 16-OCT-2003.  
PD  
XX  
XX 03-APR-2002; 2002US-00029386.  
PF  
XX  
XX 03-APR-2002; 2002US-00029386.  
PR  
XX  
XX (PENN/) PENN S G.  
PA (RANK/) RANK D R.  
PA (HANZ/) HANZEL D K.  
XX  
XX Penn SG, Rank DR, Hanzel DK;  
PI  
XX  
XX WPI; 2004-119264/12.  
DR  
XX  
XX New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
XX Claim 1; SEQ ID NO 24369; 80pp; English.  
PS  
XX  
XX The invention relates to a nucleic acid probe for measuring human gene  
XX expression, comprising any of the 27,400 fully defined nucleotide  
XX sequences in the specification, or their complements or fragments, and  
XX encoding at least 8 amino acids of any of the 6888 amino acid sequences  
XX fully defined in the specification. The probe is a single exon probe that  
XX hybridizes under high stringency conditions to a nucleic acid molecule  
XX expressed in human cells or tissues. Also included are a spatially-  
XX addressable set of single exon nucleic acid probes for measuring human  
XX gene expression (comprising a plurality of single exon nucleic acid  
XX probes cited above, where each of the plurality of probes is separately  
XX and addressably isolatable or amplifiable from the plurality), a single  
XX exon microarray for measuring human gene expression, a method of  
XX measuring human gene expression, a vector comprising the single exon  
XX probe cited above, an ORF-encoded peptide comprising at least 8  
XX contiguous amino acids of any of the above-mentioned amino acid

CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subcription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=20030194704  
XX  
XX  
SQ Sequence 174 BP; 31 A; 76 C; 36 G; 31 T; 0 U; 0 Other;  
Query Match 35.3%; Score 174; DB 12; Length 174;  
Best Local Similarity 100.0%; Pred. No. 7.2e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 44 GAGGCTGTGGCAGAAATTCCTCCGAGAGAGGCCCAAGCTGTCCCTGTGGAAGCCACGG 103  
DB 174 GAGGCTGTGGCAGAAATTCCTCCGAGAGAGGCCCAAGCTGTCCCTGTGGAAGCCACGG 115  
QY 104 CTGGCTGTGTATGAGAGAAATCTTCTCCGAGAGGCTGTGCTGTGTCAGAGAGAGCTCAG 163  
DB 114 CTGGCTGTGTATGAGAGAAATCTTCTCCGAGAGGCTGTGCTGTGTCAGAGAGAGCTCAG 55  
QY 164 GGAAGCTGTGGCAGAGTCTGTGGCGGCTTTGAGGCTGTCTGGAAGAAATCTGTCTGAG 217  
DB 54 GGAGCTGTGGCAGAGTCTGTGGCGGCTTTGAGGCTGTCTGGAAGAAATCTGTCTGAG 1  
RESULT 6  
AAL51693  
ID AAL51693 standard; cDNA; 2282 BP.  
XX  
XX AAL51693;  
AC  
XX  
XX 01-MAY-2003 (first entry)  
DT  
XX  
XX Kukulcania hibernalis spider silk protein coding sequence #1.  
DE  
XX  
XX Gene; ss; spider silk; spider silk protein; fabric; suture;  
KM medical covering; high-tech clothing; rope; reinforced plastic.  
XX  
XX Kukulcania hibernalis.  
OS  
XX  
XX WO200299082-A2.  
PN  
XX  
XX 12-DEC-2002.  
PD  
XX  
XX 06-JUN-2002; 2002WO-US018256.  
PF  
XX  
XX 06-JUN-2001; 2001US-0296184P.  
PR  
XX  
XX (UYWY-) UNIV WYOMING.  
PA  
XX  
XX Roth DA, Lewis RV;  
PI  
XX  
XX WPI; 2003-140616/13.  
DR P-PADB; AAO16494.  
XX  
XX Expressing spider silk protein in a higher plant, by contacting a plant  
PT cell with silk protein encoding a gene linked to a gene that confers  
PT resistance to selection agent, and selecting cells that survive when



Db 1726 GAGAGAGAGAGGTGTGATGTGAGAGAGAGTGTCTATGTGTGAGAGAGAGCTG 1785  
 QY 210 CTGCTG 215  
 Db 1786 GTGATG 1791

RESULT 8  
 AAQ55750  
 ID AAQ55750 standard; DNA; 1505 BP.  
 AC AAQ55750;  
 DT 25-MAR-2003 (revised)  
 DT 23-JUL-1994 (first entry)  
 DE Genomic clone G11F, includes a Brassica root specific promoter.  
 DE Plants; transformation; food crops; toxins; insects; fungal; infection;  
 KM ss.  
 OS Brassica napus.  
 FH Key Location/Qualifiers  
 FT CAAAT\_signal /tag= b 438..440  
 FT TATA\_signal /tag= b 485..488  
 FT misc\_feature /tag= c 552..554  
 FT /note= "translation start site"

MO9402619-A1.  
 03-FEB-1994.  
 12-JUL-1993; 93WO-US006541.  
 16-JUL-1992; 92US-00915246.  
 (PION-) PIONEER HI-BRED INT INC.  
 PI Byszczynski CJ, Falls L, Bellemare G, Boyvin R;  
 DR WPI; 1994-048877/06.  
 PT Brassica root-specific promoter sequence and vectors containing it - for  
 PT expression of e.g. toxins in the roots of crops susceptible to root  
 PT diseases.  
 PS Disclosure; Fig 4; 54pp; English.

The sequence shows the G11F genomic clone isolated from DNA from the  
 CC roots of Brassica napus and containing a promoter sequence upstream from  
 CC the translation start site. The promoter enhances gene expression in  
 CC roots and can be used to confer disease resistance or immunity upon a  
 CC susceptible plant root by transforming the plant with a composite gene  
 CC capable of abundant expression in the gene, comprising the promoter and  
 CC the gene of interest. This method can be used to transform plants, partic  
 CC food crops that have roots susceptible to fungal or insect diseases. The  
 CC roots of the transformed plants express a toxin or substance that  
 CC provides immunity or resistance to the fungal or insect diseases. See  
 CC also AAQ55749 and AAQ56913-5. (Updated on 25-MAR-2003 to correct PN  
 CC field.)  
 CC  
 SQ Sequence 1505 BP; 395 A; 201 C; 528 G; 381 T; 0 U; 0 Other;  
 Query Match 10.4%; Score 51.4; DB 2; Length 1505;  
 Best Local Similarity 49.8%; Pred. No. 0.0079;  
 Matches 130; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 14 GATGCGAGTCCCAAGAGCGCGAGTTTGAAGAGCTGTGCGAGAATTCCCGAGAAAGA 73

Db 852 GGTATTAGAGAGTGAAGAACGCGTGAAGTGTGTGAGCTGCATACGAGAGCGCAGAGCTGGA 911  
 QY 74 GGCCCAAGCTGTCCCTGTGTGAAGACCGAGGCTGTGCTGTGATGAGAACTCTTTCGCGA 133  
 Db 912 GCACATGTGTGAGAGAGAGCGAGAGCGTGTGTGTGAGAGCTGGCCGTCCGAT 971  
 QY 134 GGGTCTGCTCCGTGTGACGAGAGAGCTCAAGAGGCTGCAGAGTGTGGCGGCTTGAG 193  
 Db 972 GGTGTGTGATACGTGTGTGAAGAGTGTGTGTGTGTGAGAGATATGAGAGTGGCGGT 1031  
 QY 194 GCTGTGGAAGAAAGTCTGTGAGCTTCATCAGAACTGGCATTTGCAAGAGATGAAGT 253  
 Db 1032 GCAGGTGACATGAGAGTGTGTGAGAGCGGTGAAGATGAGAGCGGTGAGAGTGAAGT 1091  
 QY 254 GAATTGCGGGAAGAAATGCT 274  
 Db 1092 GCACACGCTGTGTGATACCGT 1112

RESULT 9  
 ABK63598  
 ID ABK63598 standard; cDNA; 15231 BP.  
 AC ABK63598;  
 XX  
 DT 18-JUN-2002 (first entry)  
 DE Rac sequence differentially expressed in response to a hepatotoxin #1505.  
 DE Rac  
 KM Rac; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
 KM differential expression; centrilobular necrosis; steatosis.  
 XX  
 OS Rattus norvegicus.  
 XX  
 PN WO200210453-A2.  
 PD 07-FEB-2002.  
 XX  
 PF 30-JUL-2001; 2001WO-US023872.  
 XX  
 PR 31-JUL-2000; 2000US-0222040P.  
 PR 02-NOV-2000; 2000US-0244880P.  
 PR 11-MAY-2001; 2001US-0290029P.  
 PR 15-MAY-2001; 2001US-0290645P.  
 PR 22-MAY-2001; 2001US-0292336P.  
 PR 06-JUN-2001; 2001US-0295798P.  
 PR 13-JUN-2001; 2001US-0297457P.  
 PR 19-JUN-2001; 2001US-0298884P.  
 PR 09-JUL-2001; 2001US-0303459P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Mendrick D, Porter MW, Johnson KR, Caetle AL, Elashoff MR;  
 XX  
 DR WPI; 2002-241625/29.  
 XX  
 PT Predicting toxic effects of compounds or the progression of these toxic  
 PT effects by determining the changes in gene expression in tissues or cells  
 PT exposed to the toxin and comparing these to gene expression in unexposed  
 PT tissues or cells.  
 XX  
 PS Claim 1; SEQ ID NO 1505; 239bp; English.  
 XX  
 CC The invention relates to methods for predicting toxic effects of  
 CC compounds or the progression of these toxic effects by determining the  
 CC global changes in gene expression in tissues or cells exposed to the  
 CC toxin and comparing these to gene expression in unexposed tissues or  
 CC cells. Also included are methods of predicting at least one toxic effect  
 CC of a compound or progression of a toxic effect, preferably the  
 CC hepatotoxicity of a compound, comprising detecting the level of  
 CC expression in a tissue or cell sample exposed to the compound of two or  
 CC more genes listed in the specification, where differential expression of

the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilize a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridizes to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterized by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent

Sequence 15231 BP; 3451 A; 4277 C; 4932 G; 2571 T; 0 U; 0 Other:

Query Match 10.2%; Score 50.2; DB 6; Length 15231;  
Best Local Similarity 49.8%; Pred. No. 0.029;  
Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

20 CGAGTCCCAAGAGCCGAGTTTGAAGGCTGGTGGCAGAAATCCCGAGAAAGAGGCCCA 79  
7947 CGAGCAGAGAGAGAGCCGAGTTTTCAGAGCAGAGGTGGCAAAAGCAAAACA 8006

80 GCTGTCCCTGTGAGAGCGGCTGGTGTGATGAGAAATCTTCCGAGGGTGC 139  
8007 GCTCAGAGAGAGAGCAGCGGCGAGCAGCAAGATGAGCAGAAAGCAGAGGCTGT 8066

140 TGCCGTGGTGCAGAGAGAGCTCAGGAGCTTGCAGAGTCTGTGGCGGCTTTAGGCTGCT 199  
8067 GGCAGCATGAGAGAGGCGCGAGGCGGCGAGCGTGAAGAGAGGGGTGTGAGGCGCAA 8126

200 GGAAGAAAGTCTGCTGAGCCTCATCAGAAACTGGCATTCGAGAGATGAGAAATGATTC 259  
8127 GCAAGAGAACTGCAAGCTGTGAGCAGCAGCGGAGCAGAGAGAAACTACTGGCAGA 8186

260 GGGGAGAAATGCT 274  
8187 GGAGAACCAAGAGCT 8201

RESULT 10  
ADBS9205  
ID ADBS9205 standard; DNA; 15231 BP.  
XX AC ADBS9205;  
XX DT 04-DEC-2003 (first entry)  
XX DE Toxicity-related gene, SEQ ID 4231.  
XX KM Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
XX KW drug screening; toxicity assay; ds.  
XX OS Unidentified.  
XX PN WO2003064624-A2.  
XX PD 07-AUG-2003.  
XX PF 31-JAN-2003; 2003WO-US003194.  
XX PR 31-JAN-2002; 2002US-00060087.  
XX PR 15-MAR-2002; 2002US-0364045P.  
XX PR 15-MAR-2002; 2002US-0364055P.

30-DEC-2002; 2002US-0436643P.  
(GENE-) GENE LOGIC INC.  
Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;  
WPI; 2003-689530/65.

Predicting a toxic effect of a compound, useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays, comprises preparing gene expression profile of tissue or cells exposed to the compound.

Claim 1; SEQ ID NO 4231; 1156pp; English.

The present invention relates to a method for predicting a toxic effect of a compound. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the gene expression profile to a database comprising SEQ ID 1-4925, where differential expression of the gene indicates at least one toxic effect. The method is useful for predicting at least one toxic effect of a compound, predicting hepatotoxicity or the progression of a toxic effect of a compound, identifying an agent that modulates the onset or progression of a toxic response, predicting the cellular pathways that a compound modulates in a cell, and identifying an agent that modulates at least one activity of a protein. The method and compositions of the present invention using a database of genes having liver toxin-induced differential expression, are useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 15231 BP; 3451 A; 4277 C; 4932 G; 2571 T; 0 U; 0 Other:

Query Match 10.2%; Score 50.2; DB 10; Length 15231;  
Best Local Similarity 49.8%; Pred. No. 0.029;  
Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

20 CGAGTCCCAAGAGCCGAGTTTGAAGGCTGGTGGCAGAAATCCCGAGAAAGAGGCCCA 79  
7947 CGAGCAGAGAGAGAGCCGAGTTTTCAGAGCAGAGGTGGCAAAAGCAAAACA 8006

80 GCTGTCCCTGTGAGAGCGGCTGGTGTGATGAGAAATCTTCCGAGGGTGC 139  
8007 GCTCAGAGAGAGAGCAGCGGCGAGCAGCAAGATGAGCAGAAAGCAGAGGCTGT 8066

140 TGCCGTGGTGCAGAGAGAGCTCAGGAGCTTGCAGAGTCTGTGGCGGCTTTAGGCTGCT 199  
8067 GGCAGCATGAGAGAGGCGCGAGGCGGCGAGCGTGAAGAGAGGGGTGTGAGGCGCAA 8126

200 GGAAGAAAGTCTGCTGAGCCTCATCAGAAACTGGCATTCGAGAGATGAGAAATGATTC 259  
8127 GCAAGAGAACTGCAAGCTGTGAGCAGCAGCGGAGCAGAGAGAAACTACTGGCAGA 8186

260 GGGGAGAAATGCT 274  
8187 GGAGAACCAAGAGCT 8201

RESULT 11  
ADBS3026  
ID ADBS3026 standard; DNA; 15231 BP.  
XX AC ADBS3026;  
XX DT 04-DEC-2003 (first entry)  
XX DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3568.  
XX KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
XX KW toxicity marker; toxicity progression; drug screening;  
KM primary rat hepatocyte toxicity modelling; gene; ds.











**THIS PAGE LEFT BLANK**





JOURNAL  
REFERENCE  
AUTHORS

of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3080)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanita, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

TITLE  
JOURNAL

Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
location/Qualifiers

FEATURES  
source

1..3080  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:5330429E21"  
/db\_xref="taxon:10090"  
/clone="5330429E21"  
/sex="male"  
/tissue\_type="pituitary gland"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
70..858  
/note="unnamed protein product; hypothetical Spectrin  
repeat containing protein (InterPro|IPR002017, evidence:  
InterPro)  
putative"  
/codon\_start=1  
/protein\_id="BAC27012.1"  
/db\_xref="GI:26326537"  
/translation="MKSSPEGATMVOEELRKIMESQALRLLENNMISMRNOOLR  
TEVDTGKQVPTNNIPKAGFLINPDDPIPRQHCANPLBEGHDEPDHLPQILDRDEQHL  
QAENSKLRIITKRVATAKDLRTREVKQIELEAIPEGQHLFENLRLRPDRDSNEL  
EDLRYRMWLYSKLDSGHLTESSPGLTFQFSRQKRWSPCSLLQKACRALPLQ  
LILLPLILLFLPAGEEBESCALNANFARFALMLRYNGPPT"

CDS

Query Match 38.3%; Score 188.6; DB 4; Length 3080;  
Best Local Similarity 78.0%; Pred. No. 2.6e-36;  
Matches 227; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

ORIGIN

45 AGCGTGTGCGAGATTCCCGAAGAGAGCCAGCTGTCCTCGTGGAGCGGAGGC 104  
Db 1 ACGGCTGTGTGAAATCCAGAGAAAGAGGTCCAGGTGTCCTGCTCCAGACACTGAGC 60  
Qy 105 TGCGTGTGATGAGAGAGTCTTCTCCGAGAGGTGCTGCCGTGATGACAGAGAGCTCAGG 164  
Db 61 CAGCTTGATGATGAAGAGTCTTCCCGAAGAGGCGACACATGATCCAGAGAGAGCTGAGG 120  
Qy 165 GAGCTGCGAGAGTCTGCGGCGCTTGAAGGCTGCTGGAAGAAAGTCTGCTGAGCTCATC 224  
Db 121 AAGCTGATGAGAGTCTTGGCAGGCGCTGCGGCTGTAGAGAGAAACATGCTGATCATG 180

Qy 225 AGAACTGCGATCTGACAGATGAGATGATTCGGGAGAGAAATGTTTCCACAC 284  
Db 181 AAAAAACACACACCTTCAGAGGACAGAGGTGACACGGGAGAAAGACAGGTGTTCCACAC 240  
Qy 285 AACATCCCAAGTCAGATTTCTCATCATCATCCATGATCTTATTCACAG 335  
Db 241 AACATCCCAAGGCGCGCTTCTCATCATCACTCAGACCCCATTCACAGG 291

RESULT 3  
LOCUS  
DEFINITION  
A0033373 384 bp DNA linear GSS 02-JUL-1998  
HS\_2227\_A2\_H11\_MR CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=2227 Col=22 Row=O, genomic survey  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

A0033373  
A0033373.1 GI:3285491  
GSS.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 384)  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, D., Young, D., Zhao, S., Adams, M.D. and  
Hood, L.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
COMMENT

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 2227 row: O column: 22  
Class: BAC ends  
High quality sequence stop: 384.  
Location/Qualifiers

FEATURES  
source

1..384  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=2227 Col=22 Row=O"  
/sex="male"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/note="Organ: sperm; Vector: pBelBAC11; BAC clones in  
E-Coli DH10B"

ORIGIN

Query Match 36.0%; Score 177.6; DB 9; Length 384;  
Best Local Similarity 93.8%; Pred. No. 1e-33;  
Matches 195; Conservative 0; Mismatches 12; Indels 1; Gaps 1;  
286 ACATCCCAAGTCAGATTTCTCATCATCCATGATCTTATCCAGGATCGTGCAC 345  
Db 354 ACATCCCAAGTCAGATTT-TCATCATCCAGGATCTTATTCAGGATCGTGCAC 296  
Qy 346 GCGTGAATCTGTACAGAGGCTGTGGAGAGAGGCGCAGGCCCAAGTCAAGAGTGGG 405  
Db 295 GCGTGAATCTGTACAGAGGCTGTGGAGAGAGGCGCAGGCCCAAGTCAAGAGTGGG 236  
Qy 406 TAGGGTCTCCAGACAGGCGCTCTCTGTTGGGGCAATGCTGCTTGGAGCTT 465  
Db 235 TAGGGTCTCCAGACAGGCGCTCTCTGTTGGGGCAATGCTGCTTGGAGCTT 176  
Qy 466 GGCACGTCCTGCTCATTTGAGCTGC 493

Db 175 GGCCAGCTCCTGTTTCATTGAGCTTGC 148

RESULT 4  
LOCUS AM357996 248 bp mRNA linear EST 25-APR-2001  
DEFINITION 41821 MRC 3BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AM357996  
VERSION AM357996.1 GI:6862002  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 248)  
AUTHORS Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Cassae,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,  
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,  
Chicko-Mckown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F.,  
Quackenbush,J. and Keefe,J.W.  
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
JOURNAL Genome Res. 11 (4), 626-630 (2001)  
PUBMED 11282978  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases scaled and trimmed with phred  
v0.960904.e. Vector identified by cross\_match with the -minscore 20  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCAGTCAGCAGACG  
Plate: 21 row: 1 column: 2  
Seq primer: ATTGAGTGACACTATG.  
Location/Qualifiers  
1..248  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 3BOV"  
/note="Vector: PCMV SPOT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from marrow, alveolar  
macrophage, ovary, fetal semitendinosus muscle, and fetal  
longissimus muscle."

ORIGIN

Query Match 33.1%; Score 163.2; DB 1; Length 248;  
Best Local Similarity 81.5%; Pred. No. 3,9e-30;  
Matches 189; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 3 GCCGGCCCCGGGAGTCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGCAAAATTC 62  
DB 17 GCGGGCCCATGGAGACACCCAGTCCCGGAGAGCTGAGGTTGAGAGACTGTGCTGAATTC 76  
QY 63 CCGGAGAAGAGGCCCAAGCTGTCTCCCTGTGTGAAGCCGAGGGCTGGCTGTGTGAAGGAAG 122  
DB 77 CCGGAGAAGGAACCCAGAGCTGTGAGTGAAGCCGCTTGGCCGCTGGTGAAGAGGCG 136  
QY 123 TCTTTCGCGAGGAGTGTGCGCTGTGTGACAGAGAGCTCAGGAGCTGGCAGAGTGGTG 182  
DB 137 TCTTCTCTGGAGGAGAGTGTCTGTGTCTCAAGGAGAGCTGAGAGAGCTGACCGAGTATGG 196  
QY 183 CCGGAGCTTGAAGCTGTGAAGAAAGTCTGTGAGCTCATCAGAAACTGGC 234  
DB 197 AGGCGCTTGAAGAGCTGTGAAGAGAGCTGTGAGCTCATCAGAAACCGGC 248

RESULT 5  
LOCUS AZ971121 438 bp DNA linear GSS 27-APR-2001  
DEFINITION 2M0244E01R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
clone UUGC2M0244E01 R, genomic survey sequence.  
ACCESSION AZ971121  
VERSION AZ971121.1 GI:13842348  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 438)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,W., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D.,Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 1000 Std Error: 0.00  
Plate: 0244 row: E column: 01  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 438.  
Location/Qualifiers  
1..438  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0244E01"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from The Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (GI:4732114|9b|F129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapored mouse DNA was annealed to  
adapored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

ORIGIN

Query Match 24.6%; Score 121.4; DB 9; Length 438;  
Best Local Similarity 69.8%; Pred. No. 1.4e-19;  
Matches 178; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 208 GTCTGTGAGCTCATCAGAAACTGGCATCTGACAGAGTGAAGTGTGGGGAGAA 267  
DB 114 GTCTCCCAAGTCTCATGAGAAACAGCAGCTCAGAGGACAGAGGTGACACGGGAGAA 173

Oy	268	AAATGCTTTTCCAAACAACATCCCAAGTCAGAGATTTTCATCATCATCCATGATCTTA	327
Db	174	AGCAGGTGTTACCAACAACATCCCAAGGCCCGCTTTCATCATCAACCTTAGAGACCCA	233
Oy	328	TTCCAGGCATCTGTGACGCGCTGAGTCTGTCTAGCAGGGCTGTGGAGAAAGGCGCAGGC	387
Oy	388	CCCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCGCTCCCTGTCCTGGGGCAATAT	447
Db	234	TTCCACAGGAGACAGCATGCGGTGAGTGTCTCCACAGGTTTGGAGAGA-ACGGTCAGATCC	292
Db	293	AATGTCAAGTAAAGAGGTGGGGCTTCAGACAGTTCCTCTCTGCTTGAGTGTAT	352
Oy	448	GCTCGCTCTGAGGA	462
Db	353	GCTCAGCTCATTGGGA	367
RESULT 6			
LOCUS	CN718887	596 bp	mRNA
DEFINITION	E07744B03-5 N1A Mouse four-cell-Embryo cDNA library (long) Mus		
ACCESSION	CN718887		
VERSION	CN718887.1		
KEYWORDS	GI:47488272		
SOURCE	EST.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.		
	1 (bases 1 to 596)		
	Sharov A.A., Piao Y., Matoba R., Dudekula D.B., Qian Y., VanBuren V., Palo G., Martin P.R., Stagg C.A., Bassey U.C., Wang Y., Carter M.G., Hamatani T., Albe K., Akutsu H., Sharova L., Tanaka T.S., Kimber W.L., Yoshikawa T., Jaradat S.A., Pantano S., Nagata R., Boheler K.R., Taub D., Hodess R.J., Longo D.L., Schlesinger D., Keller J., Klotz E., Keisoe G., Umezawa A., Vecovci A.L., Rossant J., Kunath T., Hogan B.L., Curci A., D'Urso M., Keisoe J., Hide W. and Ko M.S		
TITLE	Transcriptome analysis of mouse stem cells and early embryos		
JOURNAL	PLoS Biol. 1 (3), 410-419 (2003)		
COMMENT	Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@jgsun.gic.nia.nih.gov Plate: E0774 row: B column: 03 Seq primer: M3 Reverse High quality sequence stop: 596 POLYA=No.		
FEATURES			
source	location/Qualifiers		
	1..596		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6J"		
	/db_xref="niaEST:E0774B03-5"		
	/taxon="taxon:10090"		
	/clone="NIA:E0774B03 IMAGE:30903086"		
	/tissue_type="4-cell stage embryo"		
	/dev_stage="4-cell"		
	/lab_host="DH10B"		
	/clone_1ib="NIA Mouse four-cell-Embryo cDNA library (long)"		
	/notes="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://jgsun.gic.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen):		

ORIGIN	Query Match	23.9%; Score 117.6; DB 7; Length 596;
	Best Local Similarity	76.6%; Pred. No. 1,36-18;
	Matches	144; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Db	7	GCCCCGGGGGATGCCGAGTCCCAAGAGGCCAGTTTGAGAGGCTGTGGCAGAAATTCCCGG 66
Db	409	GTCGTGGGAGTGTGAGTCCCAAGAGGCTGACATCCGACACCTCTGCTCTAATAATCCAG 468
Qy	67	AGAAAGAGAGCCCTGCTGCTGCTGGGAGAGCGAGGCTGCTGCTGATGTAGAGAACTTT 126
Db	469	AGAAAGAGAGTCCAGATGTCTCTGCTCCAGACACTGGGCCACTTGTATGAAGAAGCTTT 528
Qy	127	CTCCGAGAGGTGTGCTCCGTGTGTGAGAGAGAGCTCAGGAGCTTGACAGAGTGTGGCGG 186
Db	529	CCCCAGAAAGGGGCAACATGTGTCCAGAGGAGAGCTGAGAGAACTGTAGTGAATCTTGGCAG 588
Qy	187	CCTTGAGG 194
Db	589	CCCTGCGG 596

5'-pGACTACTTCTAGATCCGAGGCGGCCCCCTTTTTTTTTTTTTTT-3' from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lona-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Plao."





TITLE Brokstein, P. and Lindquist, E.A.  
 JOURNAL DOE Joint Genome Institute Xenopus tropicalis EST project  
 COMMENT Unpublished (2004)  
 Other ESTs: JGI CAAN6348.rev  
 Contact: Lindquist, E.A., Richardson, P.  
 DOE Joint Genome Institute  
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 Tel: 925 236 5600  
 Fax: 925 236 5710  
 Email: cdna@jgi-psf.org

Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,  
 University of California, Berkeley:  
 http://tropicalis.berkeley.edu/home)  
 CDNA Library Preparation: DOE Joint Genome Institute:  
 http://www.jgi.doe.gov  
 DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
 Clone Distribution: I.M.A.G.E. Consortium/LNL:  
 http://image.lnl.gov  
 Naming Conventions: EST name is generated by the concatenation of  
 the JGI Clone id and the direction of sequencing. The suffix '.fwd'  
 indicates a forward sequencing read of the insert. It does not  
 necessarily reflect the orientation of the insert.  
 Plate: CAAN 0065 row: 9 column: 4  
 High quality sequence stop: 809.  
 Location/Qualifiers

FEATURES  
 source  
 1..858  
 /organism="Xenopus tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="IMAGE:7691934"  
 /sex="male"  
 /tissue\_type="Testes"  
 /dev\_stage="Adult"  
 /lab\_host="ElectronMAX DH10B"  
 /clone\_id="NH\_XGC\_tropT64"  
 /note="Vector: PCMVSPORT6; Site\_1: SalI; Site\_2: NotI;  
 This library was made from dt primed cDNA and cloned into  
 Invitrogen PCMVSPORT6 vector. The work was done at DOE  
 Joint Genome Institute. Poly A RNA were primed with 5'  
 GACTACTTCTAGATCCGAG CGGCGCCCTTTTCTTTT 3'. CDNA  
 were ligated to SalI adapter (5' TCACCCACCGGTCG and  
 5' GCGACGCGTGG), digested with NotI, size fractionated in  
 1.1% agarose gel electrophoresis and ligated into NotI and  
 SalI digested PCMVSPORT6 vector."

## ORIGIN

Query Match 16.9%; Score 83.2; DB 8; Length 858;  
 Best Local Similarity 58.5%; Pred. No. 6e-10;  
 Matches 145; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 6 GGGCCGGGAGTGGCGAGTCCCAAGAGCGAGTTGAGAGCGTGGGAGAAATTC 65  
 DB 64 GGGCATTTGGGATTTTGAATCCCGAGAGAAATACAGATATTTGTCTGAGCTAC 123  
 QY 66 GAGAAGAGAGCCCAAGCTCTCCCTGATGAGAGCGAGGCTGCTGTGATGAGAAAGTCT 125  
 DB 124 GAAAAAGATATCCAGCTACAAAGAGTAGAATCTCAAGTTTGTAGTAGAGAGCCCTCC 183  
 QY 126 TCTCCGAGAGGTCTGCGGTGTGAGAGAGAGCTCAGGAGCTGGCAGAGTCTGTGCGG 185  
 DB 184 TCTCCAGAGAGAGCTGCCACATTCAGACTGAGAGGAGCTTAATGCTCTCGGCTA 243  
 QY 186 GCGTTGAGAGCTGCTGAGAGAAAGTCTGCTGAGCCCTCAGAAACTGGCATCTGAGAGG 245  
 DB 244 AATCTGAAACTACTCTGGGGGCACTGGCAAGCGCTTTAAAGACAGAGATTCCAGGA 303  
 QY 246 ATGGAAGT 253  
 DB 304 CTGGCACT 311

RESULT 10  
 CNSOGARZ

LOCUS CNSOGARZ 1658 bp mRNA linear HTC 05-JUL-2005  
 DEFINITION Tetraodon nigroviridis full-length cDNA.  
 ACCESSION CR709987  
 VERSION CR709987.2 GI:56291536  
 KEYWORDS HTC; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.  
 SOURCE Tetraodon nigroviridis  
 ORGANISM Tetraodon nigroviridis

## REFERENCE

1 Jallou, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N.,  
 Maucell, E., Bonnaud, L., Fischer, C., Ozouf-Costaz, C., Bernot, A.,  
 Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C.,  
 Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N.,  
 Castellano, S., Anthouard, V., Jubin, C., Castell, V., Katinka, M.,  
 Vacherie, B., Blemont, C., Skalli, Z., Catolico, L., Poulain, J., De  
 Bernardis, V., Crnaud, C., Duprat, S., Broctier, P., Coutanceau, J.P.,  
 Gouzy, J., Parra, G., Lardier, G., Chaple, C., McKernan, K.J.,  
 McEwan, P., Bosak, S., Kellis, M., Wolf, J.N., Guigo, R., Zody, M.C.,  
 Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D.,  
 Robinson-Rechavi, M., Laudet, V., Schachter, V., Queller, F.,  
 Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissbach, J.  
 and Roest Crollins, H.  
 genome duplication in the teleost fish Tetraodon nigroviridis  
 reveals the early vertebrate proto-karyotype  
 Nature 431 (7011), 946-957 (2004)  
 15496914

## REFERENCE

## AUTHORS

## TITLE

COMMENT Direct Submission  
 Submitted (24-NOV-2004) Genoscope - Centre National de Sequencage -  
 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 (E-mail : segre@genoscope.cns.fr - Web : www.genoscope - fr)  
 On Dec 3, 2004 this sequence version replaced gi:51207896.  
 The sequences are based on single pass reads.  
 More information available at  
 http://www.genoscope.cns.fr/tetraodon.

## FEATURES

source  
 1..1658  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:99883"  
 /tissue\_type="Eggs"

## ORIGIN

Query Match 11.2%; Score 55.2; DB 4; Length 1658;  
 Best Local Similarity 52.6%; Pred. No. 0.0073;  
 Matches 120; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 4 CCGGCCCGGAGATGCCAGATCCCAAGAGCGCGAGTTGAGAGCGTGGGAGAAATTC 63  
 DB 61 CAGGAGAGTGGAGCTGAGAGGCTGTGACATGAGCTGAGAGCGCTGGTGAATTC 120  
 QY 64 CGGAGAGAGAGCCCAAGCTCTCCCTGATGAGAGCGAGGCTGCTGTGATGAGAACT 123  
 DB 121 CAGAGAGAGAGCTCCAGCTCAGACAGATGAGAGTCCAGGCTCAGAGGCTTTGAGAGA 180  
 QY 124 CTCTTCGAGAGGTGCTGCCGTGTGAGAGAGAGCTCAGAGAGCTGGCAGAGTCTGGC 183  
 DB 181 CTCTTGTGAGAGGTGCTGACAGTGTGCAAGAGCTGAAACCTGTGTGACAGAGTCTGGC 240  
 QY 184 GGGCTTGAGAGCTGCTGAGAAAGTGTGCTGAGCCCTCATCAGAACT 231  
 DB 241 TGGCACTCTATGACATCAGTCTTAATCTGCACAGAGCTGTGAACAGCT 288

RESULT 11  
 CD574999/c 695 bp mRNA linear EST 12-JUN-2003  
 LOCUS UCRP01\_02ca04.b1 Pontinus trifoliata CTV-challenged cDNA library -  
 DEFINITION Agi Pontinus trifoliata cDNA clone UCRP01\_02ca04, mRNA sequence.

ACCESSION CD574999  
 VERSION CD574999.1 GI:31670901  
 KEYWORDS EST.  
 SOURCE Poncirus trifoliata  
 ORGANISM Poncirus trifoliata  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eucosids II; Sapindales; Rutaceae; Poncirus.  
 1 (bases 1 to 695)  
 ROOSE,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,  
 WING,R. and YU,Y.  
 Development of EST Resources and New Genetic Markers for California Citrus - Poncirus trifoliata CTV-challenged phloem - AGI  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Mikeal Roose  
 Department of Botany & Plant Sciences, University of California  
 Riverside, CA, 92521-0124, USA  
 Tel: 9097874137  
 Fax: 9097874437  
 Email: mikeal.roose@ucr.edu  
 Seg primer: T7.  
 FEATURES  
 source Location/Qualifiers  
 1..695  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"  
 /cultivar="Pomeroy Op"  
 /db\_xref="taxon:37690"  
 /clone="UCRPT01.02ca04"  
 /issue\_type="Phloem"  
 /dev\_stage="10 - 30 cm shoots"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Poncirus trifoliata CTV-challenged CDNA library - AGI"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse at University of California, Riverside. The action was a open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the CTV resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate T514 over 1 year before sampling the CTV infects sweet orange, but not genotypes carrying the CTV resistance gene. Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a CDNA library was made, and 0.5 million primary lambda CDNA clones were in vivo excised to give a population of phuescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, CDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wamamaker) using the HarVest pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 10.8%; Score 53.4; DB 6; Length 695;  
 Best Local Similarity 55.0%; Pred. No. 0.017;  
 Matches 105; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
 11 GGGGAGTGGCGAGTCCCAAGAGCGCGAGTTTGAAGAGCTGGTGCAGAAATTCGCCGAGAA 70  
 416 GGGCGTTTCGACTTGGAGAGAGAGGTGCTGAGGCTTTGGCGGCGCGGTGTGCGCGA 357

71 GGAGCCCGAGCTGTCCTGCTGGTGAAGCGAGGCTGCTGTGATGAGAACTCTTCTCC 130  
 356 GGAGAGACAGCTGTGAGAGAGACTGATTTGASGCTGCTGCTGTGCTGTGTTAGGA 297  
 131 GGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 190  
 296 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 237  
 191 GAGGCTGCTGCTG 201  
 236 GCGGCTGCTGCTG 226  
 Db  
 RESULT 12  
 LOCUS CV707428/c 708 bp mRNA linear EST 03-NOV-2004  
 DEFINITION UCRPT01.0009M24.f Poncirus trifoliata CTV-challenged CDNA library - AG12 Poncirus trifoliata CDNA clone PT\_6Ea009M24, mRNA sequence.  
 ACCESSION CV707428  
 VERSION CV707428.1 GI:55289796  
 KEYWORDS EST  
 SOURCE Poncirus trifoliata  
 ORGANISM Poncirus trifoliata  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eucosids II; Sapindales; Rutaceae; Poncirus.  
 1 (bases 1 to 708)  
 ROOSE,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,  
 WING,R. and YU,Y.  
 Development of EST Resources and New Genetic Markers for California Citrus - Poncirus trifoliata CTV-challenged phloem - AG12  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Mikeal Roose  
 Department of Botany & Plant Sciences, University of California  
 Riverside, CA, 92521-0124, USA  
 Tel: 9097874137  
 Fax: 9097874437  
 Email: mikeal.roose@ucr.edu  
 Seg primer: T7.  
 FEATURES  
 source Location/Qualifiers  
 1..708  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"  
 /cultivar="Pomeroy Op"  
 /db\_xref="taxon:37690"  
 /clone="PT\_6Ea009M24"  
 /issue\_type="Phloem"  
 /dev\_stage="10 - 30 cm shoots"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Poncirus trifoliata CTV-challenged CDNA library - AG12"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse at University of California, Riverside. The action was a open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the CTV resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate T514 over 1 year before sampling the CTV infects sweet orange, but not genotypes carrying the CTV resistance gene. Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a CDNA library was made, and 0.5 million primary lambda CDNA clones were in vivo excised to give a population of phuescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, CDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona



ACCESSION CD878716  
 VERSION CD878716.1 GI:32562532  
 KEYWORDS EST  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 692)  
 REFERENCE  
 AUTHORS Genoplatte.  
 TITLE Genoplatte, a major partnership french program in plant genomics  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Genoplatte  
 Genoplatte  
 93, rue Henri Rochefort 91025 EVRY CEDEX France  
 Tel: 33 1 69 47 54 00  
 Fax: 33 1 69 47 54 10  
 This sequence has been generated in the framework of the french  
 plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>  
 and <http://genoplatte-info.infobiogen.fr>).  
 Location/Qualifiers

## FEATURES

source  
 1..692  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="recital"  
 /db\_xref="taxon:4565"  
 /clone="AZ04103H19"  
 /tissue\_type="root"  
 /clone\_1ib="AZ04"

## ORIGIN

Query Match 10.6%; Score 52.4; DB 6; Length 692;  
 Best Local Similarity 56.3%; Pred.No.0.031; Mismatches 76; Indels 0; Gaps 0;  
 Matches 98; Conservative 0;

QY	28	AAGAGCCGAGTTTGAGAGGCTGTGACAGATTCCCGAAGAGAGGCCAGCTGTCCC	87
Db	575	AGGAGGTCGATTGGGGGTGTTGTGACAGTCGAGAGAGAGGCCGATCTGGCGG	516
QY	88	TGCTGGAAGCGCAGGCGCTGCTGTATGAGAGAAGTCTTCTCCGAGAGGTCCTGCCGTGG	147
Db	515	TACGAGAGTCGAGAGAAAGTGAAGCTGGCAGCGCGGAGTCGTCGTTGATGGTTGG	456
QY	148	TGCAGAGAGAGCTCAGGAGAGCTGGCAGAGTCGTGGCGGCGCTTGAGAGCTGCTGG	201
Db	455	TGTTGAGTCGAGATGATCTGGGCCAAGCGGTGCTGCGCGAAGGATGGCGG	402

Search completed: April 27, 2006, 06:02:36  
 Job time : 3767 secs



DEVELOPMENTAL STAGE: Somatic  
TISSUE TYPE: Root  
US-07-915-246-1

Query Match 10.4%; Score 51.4; DB 2; Length 1505;  
Best Local Similarity 49.8%; Pred. No. 0.0005;  
Matches 130; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 14 GGATGCCAGATCCCAAGAGCCCGATTGAGAGCTGTGCGCAGAAATCCCGAGAAAGA 73  
DB 852 GGTATAGTAGAGTGAAGCGGTGAAGGTGTGAGCTGGAATCCAGAGCGGAGAGCTGGA 911  
QY 74 GGCCCAAGCTCTCCTGTGTGAAGCGAGGCTGTGTGTATGAGAGATCTTCTCCGA 133  
DB 912 GGACATGTGTAGAGTGAAGAGCGAGAGGTGTGTGTGAGAGCTGTGCGCAT 971  
QY 134 GGATGTGCGCGTGTGAGAGAGCTCAGAGAGCTGGCAGAGTCTGTGCGGCTTGAG 193  
DB 972 GTGTGTGATACGTGTGTGAGAAAGTCTGTCTGTGAGAGATTTGAGAGTGGCGGT 1031  
QY 194 GCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGCAGAGATGGAAT 253  
DB 1032 GCAGGTGAGATGAGAGCTGTGTGAGCGGTGAAATGAGCGGTGAGAGGTGAGGT 1091  
QY 254 GGATTCGGGGAAGAAATGAT 274  
DB 1092 GCACAGCGTGTGTGATTCGAT 1112

## RESULT 2

US-08-781-208/C  
Sequence 208, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-En  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16442 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-208

Query Match 10.1%; Score 49.8; DB 3; Length 16442;  
Best Local Similarity 49.4%; Pred. No. 0.0031;  
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 9 CCGGGGATCCGAGTCCCAAGAGCCCGATTGAGAGCTGTGCGCAGAAATTCGCGAG 68  
DB 16369 CAGAGCAGAGAGAGAGCAGAGCAGAGCAGAGCAGAGAGAGAGAGAGAGAGAG 16310  
QY 69 AAGAGCCCAAGCTGTCTGTGTGAAGCGCAGGCTGTGTGTATGAGAAATCTTCT 128  
DB 16309 CAG 16250  
QY 129 CCGAGAGGCTGTCCGTGTGAGAGAGAGCTCAGAGAGCTGCGAGATCTGTGCGGCC 188  
DB 16249 CAG 16190  
QY 189 TTGAGCTGTGAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGCAGAGATG 248  
DB 16189 AAG 16130  
QY 249 GAAGTGGATTCCGGGAGAGAA 269  
DB 16129 AAGAGAGAGAGAGAGAGAGAA 16109

## RESULT 3

US-09-618-166-208/C  
Sequence 208, Application US/09618166  
Patent No. 6583112  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-En  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/618,166  
FILING DATE: 17-Jul-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: McMaisters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 240052.419C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16442 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 208:  
US-09-618-166-208

Query Match 10.1%; Score 49.8; DB 3; Length 16442;  
Best Local Similarity 49.4%; Pred. No. 0.0031;  
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 9 CCGGGGATGCTCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGCGAGATTCCCGAG 68  
 DB 16369 CAG 16310  
 QY 69 AAGGAGCCCAAGCTGTCTCTGTGTGAAGCCGAGGCTGTGTGTGTGTGTGTGTGT 128  
 DB 16309 CAG 16250  
 QY 129 CCGGAGGCTGTCTCTGTGTGTGAAGCCGAGGCTGTGTGTGTGTGTGTGTGTGT 188  
 DB 16249 CAG 16190  
 QY 189 TTGAGCTGTGTGAAGAAAGTCTGTGAGCTTCATCAGAACTGTGCACTTGTGAGAGATG 248  
 DB 16189 AAG 16130  
 QY 249 GAAGTGATTCGGGAGAGAA 269  
 DB 16129 AAGAGAGAGAGAGAGAGAGAA 16109

RESULT 4  
 US-10-148-806-3/c  
 ; Sequence 3, Application US/10148806  
 ; Patent No. 6762042  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bai, Chang  
 ; APPLICANT: Metzger, Michael  
 ; APPLICANT: Liu, Xiaomei  
 ; TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NHL, A DNA  
 ; FILE REFERENCE: 20585P  
 ; CURRENT APPLICATION NUMBER: US/10/148, 806  
 ; CURRENT FILING DATE: 2002-06-05  
 ; PRIOR APPLICATION NUMBER: US00/33065  
 ; PRIOR FILING DATE: 2000-12-09  
 ; PRIOR APPLICATION NUMBER: 60/169,970  
 ; PRIOR FILING DATE: 1999-12-09  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 114793  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; US-10-148-806-3

Query Match 9.9%; Score 48.6; DB 3; Length 114793;  
 Best Local Similarity 54.9%; Pred. No. 0.013;  
 Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
 QY 30 GAGCGCGAGTTGAGAGGCTGTGCGAATTCGCCGAGAGAGAGAGAGAGAGAGAGAG 89  
 DB 82835 GAGGAG 82776  
 QY 90 GTGGAAGCGAG 149  
 DB 82775 GTGGAAGCGAG 82716  
 QY 150 CAG 204  
 DB 82715 GAGGAG 82661

RESULT 5  
 US-08-217-327-3/c  
 ; Sequence 3, Application US/08217327  
 ; Patent No. 5474925  
 ; GENERAL INFORMATION:  
 ; APPLICANT: John, Maliyakal E  
 ; APPLICANT: Barton, Kenneth A  
 ; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber  
 ; NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Charles and Brady  
 STREET: P.O. Box 2113  
 CITY: Madison  
 STATE: WI  
 COUNTRY: USA  
 ZIP: 53701-2113  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/217,327  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/812,233  
 FILING DATE: 19-DEC-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seay, Nicholas J  
 REGISTRATION NUMBER: 27,386  
 REFERENCE/DOCKET NUMBER: 112290831  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 608-251-5000  
 TELEFAX: 608-251-9166  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 913 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Gossypium hirsutum  
 STRAIN: Coker 312  
 TISSUE TYPE: Fiber cells  
 IMMEDIATE SOURCE:  
 CLONE: H6  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 72..716  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 72..147  
 US-08-217-327-3  
 Query Match 9.7%; Score 47.8; DB 2; Length 913;  
 Best Local Similarity 49.4%; Pred. No. 0.0039;  
 Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;  
 QY 23 GTCCAGAGAGCCGAGTTTGAAGAGGCTGTGCGAGAAATTCGCCGAGAGAGAGAGAG 82  
 DB 517 GTGCTGAG 458  
 QY 83 GTCCCTGTGAG 142  
 DB 457 GTGCGGCTGTGAG 398  
 QY 143 CGTGTGTGAG 202  
 DB 397 GTGAG 338  
 QY 203 AGAAGTGTGCTGAG 262  
 DB 337 GAG 278  
 QY 263 GAAG 273  
 DB 277 GATGAG 267

```
RESULT 6
US-07-885-970A-3/c
; Sequence 3, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
US-07-885-970A-3

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
```

```
Db      337 GGAGAGAACTGGGTGAGAGACTTGAAACTGGGGGTGAGAGAACTGCTCACTGGGGCGGT 278
Qy      263 GAAGAAATGG 273
Db      277 GATGAAGTGGG 267

RESULT 7
US-08-298-687A-3/c
; Sequence 3, Application US/08298687A
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE:
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
US-08-298-687A-3

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
```



Db 457 GGTGGGTTGCGAGTGTGGAGTGGCGGAGGTGAGAACTGTGTGAGGAGTGGCAGGA 398  
QY 143 CGTGTGCAGAGAGAGCTCAGGAGCTGGCAGACTGTGGCGGCTTTGAGGCTGTGGA 202  
Db 397 GGTGGAAGAGCTGAGGTGAGGTTGACAGAGAGAGAGAGAGAGTGTGTGACAGGA 338  
QY 203 AGAAAGTCTGCTGAGCCTCATCAGAACTGGCATCTGCAGAGATGGAATTCGGG 262  
Db 337 GGAGAGAACTGGGTGAGAGAGTGAAGCTGGGGGTGAGAGAACTGTCACTGGGGCGGT 278  
QY 263 GAGAGAAATGG 273  
Db 277 GATGAAGTGG 267

RESULT 8  
US-08-530-797-2/c  
Sequence 2, Application US/08530797  
Patent No. 5597718

## GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.  
APPLICANT: Umbeck, Paul F.  
APPLICANT: Brill, Winston J.  
TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS  
TITLE OF INVENTION: FOR ALTERED FIBER  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles and Brady  
STREET: P.O. Box 2113  
CITY: MADISON  
STATE: WISCONSIN  
COUNTRY: U.S.A.  
ZIP: 53701

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette - 3.50 inch, 800KB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word 4.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,797  
FILING DATE: 20-SEP-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/617,239  
FILING DATE: 21-NOV-90  
APPLICATION NUMBER: US 07/253,243  
FILING DATE: 04-OCT-88  
ATTORNEY/AGENT INFORMATION:  
NAME: Nicholas J. Seay  
REGISTRATION NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 1122990245  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Gossypium hirsutum  
STRAIN: Coker 312  
DEVELOPMENTAL STAGE: 15 day old fiber cells  
TISSUE TYPE: fiber cells  
IMMEDIATE SOURCE:  
LIBRARY: CKFB15A1  
CLONE: H6  
US-08-530-797-2

Query Match

9.7%; Score 47.8; DB 2; Length 913;

Best Local Similarity 49.4%; Pred. No. 0 0039;  
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 23 GTCCCAAGAGCCGAGTTTGAGAGAGCTGTGCGAATTTCCGAGAGAGAGCCAGCT 82  
Db 517 GTGGCTGAGAGAGAGCAATGAGCTGGTGGGTAGCAGAGAGAGAGATTGCTGGC 458  
QY 83 GTCCCTGTGGAAGCCAGGAGCTGGCTGTGTATGAGAACTCTTCCGAGGGTCTGC 142  
Db 457 GGTGGGTTGCGAGTGTGTGAGTGGCGGAGGTGAGAACTGTGTGAGAGTGGCAGGA 398  
QY 143 CGTGTGCAGAGAGAGCTCAGGAGCTGGCAGAGTGTGGCGGCTTTGAGGCTGTGGA 202  
Db 397 GGTGGAAGAGCTGAGGTGAGAGTTCAGAGAGAGAGAGAGAGTGTGAGTTGACAGGA 338  
QY 203 AGAAAGTCTGCTGAGCCTCATCAGAACTGGCATCTGCAGAGATGGAATTCGGG 262  
Db 337 GGAGAGAACTGGGTGAGAGAGTGAAGCTGGGGGTGAGAGAACTGTCACTGGGGCGGT 278  
QY 263 GAGAGAAATGG 273  
Db 277 GATGAAGTGG 267

RESULT 9  
US-08-298-829-3/c  
Sequence 3, Application US/08298829  
Patent No. 5620882

## GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.  
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON  
TITLE OF INVENTION: PLANTS FOR ALTERED FIBER  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
STREET: P.O. Box 2113, First Wisconsin Plaza  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: USA  
ZIP: 53701

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/298,829  
FILING DATE: 19-OCT-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/885,970  
FILING DATE: 18-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/617,239  
FILING DATE: 21-NOV-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/253,243  
FILING DATE: 04-OCT-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27,386  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 283-2478  
TELEFAX: (608) 251-5139  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

```

; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
; US-08-298-829-3

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 23 GTCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGAAATCCCGAGAGAGGCCCACT 82
DB 517 GTGGCTGAGAGAGCAAGCAATGAGCTGTGGGGGTACAGAGAGAGAGTTGCTGGC 458
QY 83 GTCCCTGTGGAAGCCGAGGCTGTGGTGTGATGAGAACTTTCTCCGAGGGGTGCTGC 142
DB 457 GTGGGGTTCAGGTGTGAGTGTGGCGGAGGTGAGAACTGTGTGAGAGAGTGGCAGGA 398
QY 143 CGTGTGACAGAGAGCTCAGAGAGCTGGCAGAGTCTGGCGGCTTGTGAGCTGTGGA 202
DB 397 GTGGAGAACTGTGAGGTGAGTTGCAGAGAGAGAGAAACAGGTGTGTGAGTTGCAGGA 338
QY 203 AGAAAGTCTGTGAGCTCATCAGAAACTGGCATCTGCAGAGATGGAATGATTCGGG 262
DB 337 GGAGGAGAACTGGTGTGAGAGAGTTGAAACTGGGGGTGTGAGAAAGTGTCACTGGGGCGGT 278
QY 263 GAAGAAAATGG 273
DB 277 GATGAAGTGGG 267

RESULT 10
; US-08-787-335-2/c
; Sequence 2, Application US/08787335
; Patent No. 5981834
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O BOX 2113
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800kb storage
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,335
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,797
; FILING DATE:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
; US-08-787-335-2

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 23 GTCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGAAATCCCGAGAGAGGCCCACT 82
DB 517 GTGGCTGAGAGAGCAAGCAATGAGCTGTGGGGGTGACAGAGAGAGATTGCTGGC 458
QY 83 GTCCCTGTGGAAGCCGAGGCTGTGTGATGAGAAAGTCTTCTCCGAGGGGTGCTGC 142
DB 457 GTGGGGTTCAGGTGTGAGTGTGGCGGAGGTGAGAAAGTGTGTGAGAGATGGCAGGA 398
QY 143 CGTGTGACAGAGAGCTCAGAGAGCTGGCAGAGTCTGGCGGCTTGTGAGCTGTGGA 202
DB 397 GTGGAGAACTGTGAGGTGAGTTGCAGAGAGAGAGAAACAGGTGTGTGAGTTGCAGGA 338
QY 203 AGAAAGTCTGTGAGCTCATCAGAAACTGGCATCTGCAGAGATGGAATGATTCGGG 262
DB 337 GGAGGAGAACTGGTGTGAGAGAGTTGAAACTGGGGGTGTGAGAAAGTGTCACTGGGGCGGT 278
QY 263 GAAGAAAATGG 273
DB 277 GATGAAGTGGG 267

RESULT 11
; US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
; US-09-249-585A-2

Query Match          9.7%; Score 47.8; DB 3; Length 1926;
Best Local Similarity 52.2%; Pred. No. 0.005;
Matches 106; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 2 GGCCGGCCCGGGGATGCCGAGTCCCAAGAGCCGAGTTTGAAGAGCTGTGTGCAGAAAT 61
DB 364 GGAGGGGCAAGAGAGAGAGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGG 423

```





Thu Apr 27 17:45:33 2006

Search completed: April 27, 2006, 03:30:35  
Job time : 163 secs

us-10-071-510a-16.rni

**THIS PAGE LEFT BLANK**



QY	241	AGAGATGGAAGTGGATTGGGGAGAAATGGTTTACCAACAATCCCAAATGCAG	300
Db	241	AGAGATGGAAGTGGATTGGGGAGAAATGGTTTACCAACAACAATCCCAAATGCAG	300
QY	301	GATTTCTCATCAATCCCATGATCTTATTTCCAGGCATGTCGACGCGTAGTCTGTCTA	360
Db	301	GATTTCTCATCAATCCCATGATCTTATTTCCAGGCATGTCGACCGCGTAGTCTGTCTA	360
QY	361	GCAGGGCTGTGGAGAGAGGGGCCAGGCCCAAGAGTGGAGTGAAGGGGTCTCCAGCA	420
Db	361	GCAGGGCTGTGGAGAGAGGGGCCAGGCCCAAGAGTGGAGTGAAGGGGTCTCCAGCA	420
QY	421	CAGGCCCTCCCTGTCTGGGGCAACAATGCTCTGTGAGAGACTTGGCCACGTCCTGTCT	480
Db	421	CAGGCCCTCCCTGTCTGGGGCAACAATGCTCTGTGAGAGACTTGGCCACGTCCTGTCT	480
QY	481	CATTGAGCCTGC	493
Db	481	CATTGAGCCTGC	493

```

RESULT 2
US-10-450-763-19758
: Sequence 19758, Application US/10450763
: Publication No. US20050196754A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc
: TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
: FILE REFERENCE: 790CIP3/US
: CURRENT APPLICATION NUMBER: US/10/450.763
: PRIOR FILING DATE: 2003-06-11
: PRIOR APPLICATION NUMBER: PCT/US01/08631
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: 09/649,167
: PRIOR FILING DATE: 2000-08-23
: NUMBER OF SEQ ID NOS: 60736
: SOFTWARE: Clustom
: SEQ ID NO 19758
: LENGTH: 4083
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SIMILAR
: LOCATION: (3462)..(3854)
: OTHER INFORMATION: 62% homologous to Plasmodium falciparum histidine-rich
: OTHER INFORMATION: protein, accession number X01469, Smith-Waterman Score=535
US-10-450-763-19758

```

Query Match	70.5%	Score 347.6	DB 9	Length 4083
Beet Local Similarity	98.9%	Pred. 3.4e-93		
Matches 350	Conservative	0	Mismatches 4	Indels 0
			Gaps	0
QY	2	GGCCGGCCCCGGGGGAGATGCCAGTCCCAAGAGCCGAGTTTGAGAGCGCTGTGGCAGATT	61	
Db	2796	GGCCGGCCCCGGGGGAGATGCCAGTCCCAAGAGGCCAGTTTGAGAGCGCTGTGGCAGATT	2855	
QY	62	CCCGGAGAGGAGGCGCCAGCTGTCTCTGTGTGAGACGCGAGGGCTGGTGTGATGAGAA	121	
Db	2856	CCCGGAGAGGAGGCGCCAGCTGTCTCTGTGTGAGACGCGAGGGCTGGTGTGATGAGAA	2915	
QY	122	GTCTTCTTCGGAAGGGTGCTGCCGTGTGTGACAGAGGAGCTCAGGGAGCTGGCAGAGTCTGT	181	
Db	2916	GTCTTCTTCGGAAGGGTGCTGCCGTGTGTGACAGAGGAGCTCAGGGAGCTGGCAGAGTCTGT	2975	
QY	182	GGGGGCTTTGAGGCTGCTGCGAAGAAAGTCTGTAGCCTCATCAAAACTGGCATCTCA	241	
Db	2976	GGGGGCTTTGAGGCTGCTGCGAAGAAAGTCTGTAGCCTCATCAAAACTGGCATCTCA	3035	
QY	242	GAGATGGAAGTGGATTGCGGGGAAATAATGGTTTCAACCAACATCTCCAAATCTAG	301	
Db	3036	GAGATGGAAGTGGATTGCGGGGAAATAATGGTTTCAACCAACATCTCCAAATCTAG	3099	

```

Oy      302 ATTTCATCAATCCCATGATTCCTATTTCCAGGCATCGTCGACGGGTGAGTCT 355
        |||||
Db      3096 ATTTCTCATCAATCCCATGATTCCTATTTCCAGGCATCGTCGACGGCAAAATCT 3145

```

```

RESULT 3
US-10-029-386-10669/c
; Sequence 10669, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENOTYPING
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10669
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR14.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P10476, EVALUE 5.70e-01
; OTHER INFORMATION: NT HIT: AJ277661.1, EVALUE 5.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: AW842384.1, EVALUE 3.00e-03
US-10-029-386-10669

```

Query Match	35.8%	Score 176.4	DB 6	Length 532
Best Local Similarity	99.4%	Pred. No. 2.9e-42		
Matches 177	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	40	TTGAGAGGCTGGTGGCAGAAATTC	CCGAGAGAGAGGCCACGCTTC	CCCTGGTGAAGCGC 99
Db	425	TTTTCAGAGGCTGGTGGCAGAAATTC	CCGAGAGAGAGGCCACGCTTC	CCCTGGTGAAGCGC 366
QY	100	AGGGCTGGCTGGTGATGAGAGAGTCTT	CTCCGAGAGGCTGCTGCCGTGGTCAGAGCAGC	159
Db	365	AGGGCTGGCTGGTGATGAGAGAGTCTT	CTCCGAGAGGCTGCTGCCGTGGTCAGAGCAGC	306
QY	160	TCAGGAGAGCTGGCAGAGTCTGGGGGGGCTT	GAGAGCTGGTCGAGAGAAAGTCTGGCAG	217
Db	305	TCAGGAGAGCTGGCAGAGTCTGGGGGGGCTT	GAGAGCTGGTCGAGAGAAAGTCTGGCAG	248

```

? RESULT 4
? US-10-029-386-24369/c
? Sequence 24369, Application US/10029386
? Publication No. US20030194704A1
? GENERAL INFORMATION:
? APPLICANT: Penn, Sharon G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
? TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
? FILE REFERENCE: AEMICA-X-2
? CURRENT APPLICATION NUMBER: US/10/029,386
? CURRENT FILING DATE: 2001-12-20
? NUMBER OF SEQ ID NOS: 34288
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 24369
? LENGTH: 174
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO CHR14.3
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1

```



```
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q99372, EVALU 6.60e-01
; OTHER INFORMATION: EST_HUMAN HIT: B125532.1, EVALU 5.70e-02
US-10-029-386-24369

Query Match      35.3%; Score 174; DB 6; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.1e-41;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GAGGCTGGGCGCAATTCGCCGAGAGAGGCGCCAGCTCTCCCTGTGAGGCGAGG 103
DB 174 GAGGCTGGGCGCAATTCGCCGAGAGAGGCGCCAGCTCTCCCTGTGAGGCGAGG 115
QY 104 CTGGCTGGTATGAGAGAGTCTTCTCCGAGAGGCTGCTCCGCTGAGAGAGAGCTCAG 163
DB 114 CTGGCTGGTATGAGAGAGTCTTCTCCGAGAGGCTGCTCCGCTGAGAGAGAGCTCAG 55
QY 164 GGAGCTGGCAGAGTCTGTCGGCGGCTTGAAGCTGTGAAAGAAAGTCTGCTGAG 217
DB 54 GGAGCTGGCAGAGTCTGTCGGCGGCTTGAAGCTGTGAAAGAAAGTCTGCTGAG 1

RESULT 5
US-10-437-963-93121/c
; Sequence 93121, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 93121
; LENGTH: 6302
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_91537C.1
US-10-437-963-93121

Query Match      11.6%; Score 57; DB 7; Length 6302;
Best Local Similarity 55.2%; Pred. No. 2.4e-06;
Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1 CGGCGGCGCCGCGGATGCCAGTCCCAAGAGCGCGAGTTTGAAGAGCTGTGCGAGAT 60
DB 5109 CGAGCGAGCTTGGGAGACCTTGTCTTGGAGGTGCAAGCGGACATGTGAGGCGGAGG 5050
QY 61 TCCCGAGAGAGAGGCCAGCTGTCTCCGTGTGGAAGCCGAGGCGCTGTGTGATGAGA 120
DB 5049 AGCCGAGAGAGATGTGTGAGAGAGAGGTGGCGAGAGACTGAGAGATGTGTGAGAGG 4990
QY 121 AGCTTCTCCGAGAGGTCTGCTCCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGTGT 180
DB 4989 AGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4930
QY 181 GGCGGCGCTTGAAGGCTGTG 201
DB 4929 TGCACGAGATGCCGCTTGTGTG 4909

RESULT 6
US-10-479-638-7
; Sequence 7, Application US/10479638

; Publication No. US20040210956A1
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; APPLICANT: The University of Wyoming
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: WYO 02-0004US
; CURRENT APPLICATION NUMBER: US/10/479,638
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/US02/18256
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2282
; TYPE: DNA
; ORGANISM: Kukulcania hibernalis
US-10-479-638-7

Query Match      11.5%; Score 56.8; DB 8; Length 2282;
Best Local Similarity 52.5%; Pred. No. 2.2e-06;
Matches 124; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 38 GTTTGAGAGCTGTGTGCGAGAAATTCGCCGAGAGAGGCGCCAGCTTCCCTGTGAGAGC 97
DB 1156 GCTGACAGCGCTGCTGTGAGAGAGGCTTCTGGAAGCGAGAGCAGAGAGCGCTACATCTGC 1215
QY 98 GCAGGCTGCTGTGTATGAGAGAGTCTTCTCCGAGAGGCTGTGCGCTGTGAGAGAGA 157
DB 1216 GAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1275
QY 158 GCTCAGGAGCTGTGCGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 217
DB 1276 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1335
QY 218 CTTGATCAGAACTGCGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
DB 1336 GGCAGGCTGTATGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1391

RESULT 7
US-10-437-963-16557/c
; Sequence 16557, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 16557
; LENGTH: 6402
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_22294C.1
US-10-437-963-16557

Query Match      11.4%; Score 56; DB 7; Length 6402;
Best Local Similarity 55.7%; Pred. No. 4.9e-06;
Matches 107; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
```

```
QY 10 CGGGGATGCGGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGAGCAAAATTCGCGAGA 69
DB 5146 CGAGAGAGCGCGAGCTCGAGAGATGTGCAGAGAGGTGGCAGAGAGCCGCGAGG 5087
QY 70 AGAGAGCCGAGCTGTCTCTGTGGAAGCCGAGGCTGTGCTGTGATGAGAAATCTTCTC 129
DB 5086 AGATGTGACAGAGAGAGGTGGCGGAGAGAGCCGAGAGATAGTGCAGAGAGAGTGGCGG 5027
QY 130 CGAGAGGTGTCTCCGCTGTGTGTCAGAGAGAGCTCAGAGAGCTGTGCGAGAGTGTGCGGCGCT 189
DB 5026 AGAGAGCCGAGAGATGTGTGTCAGAGAGAGAGCGCGAGCCGAGAGAGATGTGTGTCAGAGA 4967
QY 190 TGAGGCTGTG 201
DB 4966 CGCCGCTGTG 4955
```

## RESULT 8

```
US-10-437-963-16601/C
; Sequence 16601, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 16601
; LENGTH: 6296
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_22334C.1
US-10-437-963-16601
```

```
Query Match 10.9%; Score 53.6; DB 7; Length 6296;
Best Local Similarity 54.6%; Pred. No. 2.5e-05;
Matches 107; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
```

```
QY 1 CGGCGGCGCGGCGGAGTCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGCGAGAAAT 60
DB 5125 CGGTGAGCTTGGCAAGACCTTGTCTTGGAGGTGCGAGCGGCGACGCTGAGCGGAGG 5066
QY 61 TCCCGAGAGAGAGAGCCGAGCTGTCTCTGTGGAAGCGAGGCTGTGCTGTGATGAGAA 120
DB 5065 ARCTGAGAGAGATGTGTGACGCGGAGGTGCGAGAGAGCGCGAGGTGATGTGTCAGAGG 5006
QY 121 AGTCTTCTCCGAGAGGTGTGCTGCGCTGTGTGTCAGAGAGAGCTCAGAGAGCTGTGCGAGTGT 180
DB 5005 ACCTGTGTGAGAGAACTGAGAGAGATGTGTGTCAGAGAGAGATGTGCGCGAGAGATGG 4946
QY 181 GCGCGGCTGTGAGGCT 196
DB 4945 TGCACGAATGCTCT 4930
```

## RESULT 9

```
US-10-437-963-3998/C
; Sequence 3998, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 3998
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_10920C.1
US-10-437-963-3998
```

```
Query Match 10.8%; Score 53; DB 7; Length 458;
Best Local Similarity 52.0%; Pred. No. 2e-05;
Matches 142; Conservative 0; Mismatches 130; Indels 1; Gaps 1;
```

```
QY 1 CGGCGGCGCGGCGGAGTCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGCGAGAAAT 60
DB 373 CGGCGGAGAGAGAGAGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
QY 61 TCCCGAGAGAGAGAGCCGAGCTGTCTCTGTGGAAGCCGAGAGCTGTGCTGTGATGAGAA 120
DB 313 AGAGAGAGAGAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
QY 121 AGTCTTCTCCGAGAGGTGTGCTGCTGTGTGTCAGAGAGAGCTCAGAGAGCTGTGCGAGATC-G 179
DB 253 AGCGGAGCGGAGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
QY 180 TGGCGGCGCTTGTGAGGCTGTGTGGAAGAACTGTGAGCTTCATCAAACTGGCATCTG 239
DB 193 TGGCGGTGCGCGCGCGCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 134
QY 240 CAGAGATGAGAGTGTGATTCGCGGAGAAATATG 272
DB 133 CGAGAGAGAGATGTGTGTGTGCGGAGCAAAATTTG 101
```

## RESULT 10

```
US-10-437-963-56215/C
; Sequence 56215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 56215
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_58148C.1
US-10-437-963-56215
```

```
Query Match 10.8%; Score 53; DB 7; Length 1023;
Best Local Similarity 55.0%; Pred. No. 2.4e-05;
```

Matches 104; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 13 GGGATGCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGCGAGAAATTCGCGAGAGG 72  
 Db 220 GGGAGGTGCGAGGCGACACCGGTGAGTGAAGCGCGAGAGCGCGAGAGCCGAGAGAGA 161  
 QY 73 AGGCCAGCTGTCTCTGTGTGAAGCCGAGGCTGTGCTGTGATGAGAGATCTTCTCCG 132  
 Db 160 TGGTGACGAGAGAGTGTGTAGAGAGCCGAGAGATGTGTCAGAGAGAGTGTGACAGAG 101  
 QY 133 AGGATGCTCCGCTGTGTCAGAGAGCTCAAGGAGCTGGAGAGTGTGCGGCGCTTGA 192  
 Db 100 AGCCGAGAGATGTGTGACAGAGAGCGCGAGCCGAGAGATGTGTGACAGAGCC 41  
 QY 193 GGGTCTGG 201  
 Db 40 CGCTTGTG 32

RESULT 11  
 US-10-437-963-93061/C  
 ; Sequence 93061, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 93061  
 ; LENGTH: 6112  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_91482C.1  
 US-10-437-963-93061

Query Match 10.8%; Score 53; DB 7; Length 6112;  
 Best Local Similarity 55.8%; Pred. No. 3.8e-05;  
 Matches 101; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 21 GAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGCGAGAAATTCGCGAGAGAGCCGAG 80  
 Db 4855 GAGGTGCGAGGCGGACACGCTGAGCGGAGAGCGCGAGAGCCGAGAGATGTGTGAG 4796  
 QY 81 CTGTCTCTGTGTGAAGCCGAGGCTGTGCTGTGATGAGAGATCTTCTCCGAGGCTGCT 140  
 Db 4795 CAGAGAGTGTGCGAGAGACCCAGAGAGATGTGTGCGAGAGTGTGCGGAGAGCCAGAG 4736  
 QY 141 GCCGTGTGTCAGAGAGAGCTCAAGGAGCTGTGCGAGAGTGTGCGGCGCTTGAAGCTGTG 200  
 Db 4735 GAGATGTGTGTCAGAGAGATGTGCGAGCGCGAGAGAGATGTGTGTCAGAAATGCCGCTTGTG 4676  
 QY 201 G 201  
 Db 4675 G 4675

RESULT 12  
 US-10-437-963-16524/C  
 ; Sequence 16524, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 16524  
 ; LENGTH: 6417  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_22264C.1  
 US-10-437-963-16524

Query Match 10.8%; Score 53; DB 7; Length 6417;  
 Best Local Similarity 55.0%; Pred. No. 3.8e-05;  
 Matches 104; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 13 GGGATGCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGCGAGAAATTCGCGAGAGG 72  
 Db 5115 GGGAGGTGCGAGGCGACACGCTGAGTGTGAGAGCGCGAGAGAGCCGAGAGAGA 5056  
 QY 73 AGGCCAGCTGTCTCTGTGTGAAGCCGAGGCTGTGCTGTGATGAGAGATCTTCTCCG 132  
 Db 5055 TGGTGACGAGAGAGTGTGTGAGAGACCCGAGAGAGATGTGTGCGAGAGAGTGTGCGAGAG 4996  
 QY 133 AGGATGCTCCGCTGTGTCAGAGAGCTCAAGGAGCTGGAGAGTGTGCGGCGCTTGA 192  
 Db 4995 AGCCGAGAGAGATGTGTGACAGAGAGCGCGAGCCGAGAGAGATGTGTGACAGAGCGC 4936  
 QY 193 GGGTCTGG 201  
 Db 4935 CGCTTGTG 4927

RESULT 13  
 US-10-437-963-16612/C  
 ; Sequence 16612, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 16612  
 ; LENGTH: 6337  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_22344C.1  
 US-10-437-963-16612

Query Match 10.6%; Score 52.4; DB 7; Length 6337;  
 Best Local Similarity 54.0%; Pred. No. 5.8e-05;  
 Matches 107; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```
QY      1 CGGCCGCCCCGGGGGATGCCGAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAAT 60
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      5166 CGGTGAGCTTGGCCAAACACCTTGTCTTGGAGAGTGCAGGGCGACACGCTGAGAGCCGAGG 5107
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      61 TCCCGAGAGAGAGAGCCCACTGTCCCTGTGGAAGCGCAGGGCTGTGGTGTATGAGA 120
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      5106 AGCCGAGGAGATGTGTGCACAGAGAGGTGCGGAGGAGCCGAGAGATGTGTGCAGAGG 5047
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      121 AGCTCTTCTCCGAGAGGCTGTCCCTGTGTGCAGAGAGCTTCAGAGAGCTGTGCAGAGTCT 180
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      5046 ACCTGTGCAGAGAACCGAGAGAGATGTGTGCAGAGAGATGTGTGCAGAGAGATGCGCG 4987
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      181 GCGCGGCTTGAAGCTGC 198
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      4986 AGCCGAGAGAGATGTGTC 4969
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

## RESULT 14

```
US-10-437-963-69317/c
; Sequence 69317, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69317
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_69998C.1
US-10-437-963-69317
```

```
Query Match      10.3%; Score 50.6; DB 7; Length 2139;
Best Local Similarity 53.2%; Pred. No. 0.00015;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY      1 CGGCCGCCCCGGGGGATGCCGAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAAT 60
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      1062 CGGTGAGCTTGGCCAAACACCTTGTCTTGGAGAGTGCAGGGCGACACGCTGAGAGTGAAG 1003
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      61 TCCCGAGAGAGAGAGCCCACTGTCCCTGTGGAAGCGCAGGGCTGTGGTGTATGAGA 120
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      1002 AGCCGAGGAGATGTGTGCACAGAGAGTGTGCGGAGGAGCCGAGAGATGTGTGCAGAGATG 943
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      121 AGCTCTTCTCCGAGAGGCTGTCCCTGTGTGCAGAGAGCTTCAGAGAGCTGTGCAGAGTCT 180
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      942 ACCTGTGCAGAGAACCGAGAGAGATGTGTGAGAGAGATGTGTGCGAGCCGAGAGAGATG 883
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      181 GCGCGGCTTGAAGCTGTG 201
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      882 TGCACAGAGAGCGCGCTTGTG 862
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

## RESULT 15

```
US-10-437-963-69371/c
; Sequence 69371, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69371
; LENGTH: 4701
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70045C.1
US-10-437-963-69371
```

```
Query Match      10.3%; Score 50.6; DB 7; Length 4701;
Best Local Similarity 53.2%; Pred. No. 0.00018;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY      1 CGGCCGCCCCGGGGGATGCCGAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAAT 60
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      3470 CGGTGAGCTTGGCCAAACACCTTGTCTTGGAGAGTGCAGGGCGACACGCTGAGAGCCGAGG 3411
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      61 TCCCGAGAGAGAGAGCCCACTGTCCCTGTGGAAGCGCAGGGCTGTGGTGTATGAGA 120
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      3410 AGCCAGAGAGATGTGTGCAGTGTGAGAGTGTGCGGAGAGCCGAGAGATGTGTGCAGAGAG 3351
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      121 AGCTCTTCTCCGAGAGGCTGTCCCTGTGTGCAGAGAGCTTCAGAGAGCTGTGCAGAGTCT 180
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      3350 ACCTGTGCAGAGAACCGAGAGAGATGTGTGCAGAGAGATGTGTGCGAGACCGAGAGAGATG 3291
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      181 GCGCGGCTTGAAGCTGTG 201
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      3290 TGCACAGAGAGCGCGCTTGTG 3270
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```
Search completed: April 27, 2006, 05:37:22
Job time : 610 secs
```

November 2005

Published\_Applications Nucleic Acid and Published\_Applications Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases; older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New).

Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).

**THIS PAGE LEFT BLANK**